PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/31, C07K 14/35, 16/12, C12Q 1/68, C12N 15/62, G01N 33/53

A2

(11) International Publication Number: **WO 98/16645**

(43) International Publication Date: 23 April 1998 (23.04.98)

(21) International Application Number: PCT/US97/18214

(22) International Filing Date: 7 October 1997 (07.10.97)

(30) Priority Data:

08/729,622	11 October 1996 (11.10.96)	US
08/818,111	13 March 1997 (13.03.97)	US

(71) Applicant: CORIXA CORPORATION [US/US]; 1124 Columbia Street, Seattle, WA 98104 (US).

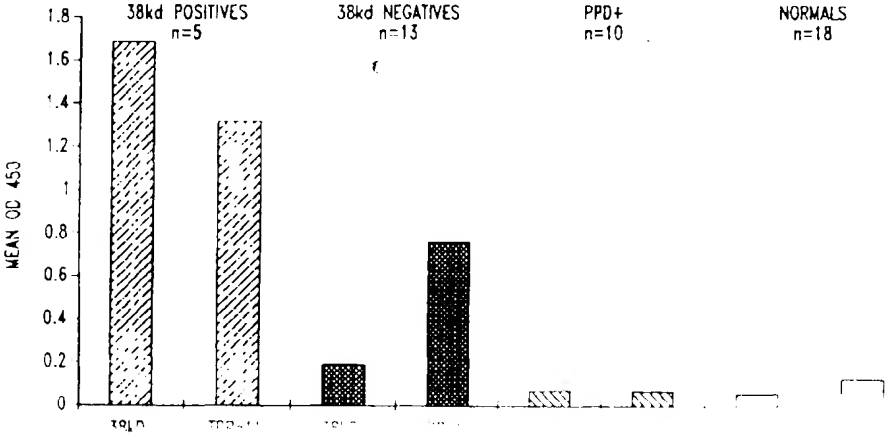
(72) Inventors: REED, Steven, G.; 2843 – 122nd Place N.E., Bellevue, WA 98005 (US). SKEIKY, Yasir, A., W.; 8327 – 25th Avenue N.W., Seattle, WA 98107 (US). DILLON, Davin, C.; 21607 N.E. 24th Street, Redmond, WA 98053 (US). CAMPOS-NETO, Antonio; 9308 Midship Court N.E., Bainbridge Island, WA 98021 (US). HOUGHTON, Raymond; 2636 – 242nd Place S.E., Bothell, WA 98021 (US). VEDVICK, Thomas, S.; 124 South 300th Place, Federal Way, WA 98003 (US). TWARDZIK, Daniel, R.; 10195 South Beach Drive, Bainbridge Island, WA 98110 (US). LODES, Michael, J.; 9223 – 36th Avenue S.W., Seattle, WA 98126 (US).

(74) Agents: MAKI, David, J. et al., Seed and Berry LLP, 6300 Columbia Center, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).

(81) Designated States: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel				
BY	Belarus						

COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

TECHNICAL FIELD

The present invention relates generally to the detection of *Mycobacterium tuberculosis* infection. The invention is more particularly related to polypeptides comprising a *Mycobacterium tuberculosis* antigen, or a portion or other variant thereof, and the use of such polypeptides for the serodiagnosis of *Mycobacterium tuberculosis* infection.

BACKGROUND OF THE INVENTION

Tuberculosis is a chronic, infectious disease, that is generally caused by infection with *Mycobacterium tuberculosis*. It is a major disease in developing countries, as well as an increasing problem in developed areas of the world, with about 8 million new cases and 3 million deaths each year. Although the infection may be asymptomatic for a considerable period of time, the disease is most commonly manifested as an acute inflammation of the lungs, resulting in fever and a nonproductive cough. If left untreated, serious complications and death typically result.

Although tuberculosis can generally be controlled using extended antibiotic therapy, such treatment is not sufficient to prevent the spread of the disease. Infected individuals may be asymptomatic, but contagious, for some time. In addition, although compliance with the treatment regimen is critical, patient behavior is difficult to monitor. Some patients do not complete the course of treatment, which can lead to ineffective treatment and the development of drug resistance.

Inhibiting the spread of tuberculosis will require effective vaccination and accurate, early diagnosis of the disease. Currently, vaccination with live bacteria is the most efficient method for inducing protective immunity. The most common Mycobacterium for this purpose is Bacillus Calmette Guerin (BCG).

site by 48-72 hours after injection, which indicates exposure to Mycobacterial antigens. Sensitivity and specificity have, however, been a problem with this test, and individuals vaccinated with BCG cannot be distinguished from infected individuals.

While macrophages have been shown to act as the principal effectors of *M. tuberculosis* immunity, T cells are the predominant inducers of such immunity. The essential role of T cells in protection against *M. tuberculosis* infection is illustrated by the frequent occurrence of *M. tuberculosis* in AIDS patients, due to the depletion of CD4 T cells associated with human immunodeficiency virus (HIV) infection. Mycobacterium-reactive CD4 T cells have been shown to be potent producers of gamma-interferon (IFN-γ), which, in turn, has been shown to trigger the anti-mycobacterial effects of macrophages in mice. While the role of IFN-γ in humans is less clear, studies have shown that 1,25-dihydroxy-vitamin D3, either alone or in combination with IFN-γ or tumor necrosis factor-alpha, activates human macrophages to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN-γ stimulates human macrophages to make 1,25-dihydroxy-vitamin D3. Similarly, IL-12 has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection *see* Chan and Kaufmann, in *Tuberculosis: Pathogenesis, Protection and Control*, Bloom (ed.), ASM Press, Washington, DC, 1994.

Accordingly, there is a need in the art for improved diagnostic methods for detecting tuberculosis. The present invention fulfills this need and further provides other related advantages.

SUMMARY OF THE INVENTION

Briefly stated, the present invention provides compositions and methods for diagnosing tuberculosis. In one aspect, polypeptides are provided comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of such an antigen that differs only

(b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);

(c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 117);

(d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);

(e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);

(f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 120);

(g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser (SEQ ID NO: 121);

(h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);

(i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn (SEQ ID NO: 123);

(j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID NO: 129)

(k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID NO: 130) or

(l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID NO: 131)

wherein Xaa may be any amino acid

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132) or

(n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124)

wherein Xaa may be any amino acid.

In another embodiment, the soluble *M. tuberculosis* antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96 or a complement thereof under moderately stringent conditions.

In a related aspect, the polypeptides comprise an antigenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, wherein the antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 26-51, 133, 134, 158-178 and 196, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 26-51, 133, 134, 158-178 and 196 or a complement thereof under moderately stringent conditions.

In related aspects, DNA sequences encoding the above polypeptides, recombinant expression vectors comprising these DNA sequences and host cells transformed or transfected with such expression vectors are also provided.

In another aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, an inventive polypeptide and a known *M. tuberculosis* antigen.

In further aspects of the subject invention, methods and diagnostic kits are provided for detecting tuberculosis in a patient. The methods comprise: (a) contacting a [illegible] with [illegible] first and urine. The diagnostic kit [illegible]

The present invention also provides methods for detecting *M. tuberculosis* infection comprising: (a) obtaining a biological sample from a patient; (b) contacting the sample with at least one oligonucleotide primer in a polymerase chain reaction, the oligonucleotide primer being specific for a DNA sequence encoding the above polypeptides; and (c) detecting in the sample a DNA sequence that amplifies in the presence of the first and second oligonucleotide primers. In one embodiment, the oligonucleotide primer comprises at least about 10 contiguous nucleotides of such a DNA sequence.

In a further aspect, the present invention provides a method for detecting *M. tuberculosis* infection in a patient comprising: (a) obtaining a biological sample from the patient; (b) contacting the sample with an oligonucleotide probe specific for a DNA sequence encoding the above polypeptides; and (c) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe. In one embodiment, the oligonucleotide probe comprises at least about 15 contiguous nucleotides of such a DNA sequence.

In yet another aspect, the present invention provides antibodies, both polyclonal and monoclonal, that bind to the polypeptides described above, as well as methods for their use in the detection of *M. tuberculosis* infection.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE IDENTIFIERS

Figure 1A and B illustrate the stimulation of proliferation and interferon γ production in T cells derived from a first and a second *M. tuberculosis*-immune donor, respectively, by the 14 Kd, 20 Kd and 26 Kd antigens described in Example 1.

Figure 3A illustrates the stimulation of proliferation in a TbH-9-specific T cell clone by secretory *M. tuberculosis* proteins, recombinant TbH-9 and a control antigen, TbRa11.

Figure 3B illustrates the stimulation of interferon-γ production in a TbH-9-specific T cell clone by secretory *M. tuberculosis* proteins, PPD and recombinant TbH-9.

Figure 4 illustrates the reactivity of two representative polypeptides with sera from *M. tuberculosis*-infected and uninfected individuals, as compared to the reactivity of bacterial lysate.

Figure 5 shows the reactivity of four representative polypeptides with sera from *M. tuberculosis*-infected and uninfected individuals, as compared to the reactivity of the 38 kD antigen.

Figure 6 shows the reactivity of recombinant 38 kD and TbRa11 antigens with sera from *M. tuberculosis* patients, PPD positive donors and normal donors.

Figure 7 shows the reactivity of the antigen TbRa2A with 38 kD negative sera.

Figure 8 shows the reactivity of the antigen of SEQ ID NO: 60 with sera from *M. tuberculosis* patients and normal donors.

Figure 9 illustrates the reactivity of the recombinant antigen TbH-29 (SEQ ID NO: 137) with sera from *M. tuberculosis* patients, PPD positive donors and normal donors as determined by indirect ELISA

Figure 10 illustrates the reactivity of the recombinant antigen TbH-33 (SEQ ID NO: 140) with sera from *M. tuberculosis* patients and from normal donors, and with a pool of sera from *M. tuberculosis* patients, as determined both by direct and indirect ELISA

Figure 11 illustrates the reactivity of increasing concentrations of the recombinant antigen TbH-33 (SEQ ID NO: 140) with sera from M. tuberculosis patients and from normal donors as determined by ELISA.

SEQ. ID NO. 5 is the DNA sequence of TbRa13.

SEQ. ID NO. 6 is the DNA sequence of TbRa16.

SEQ. ID NO. 7 is the DNA sequence of TbRa17.

SEQ. ID NO. 8 is the DNA sequence of TbRa18.

SEQ. ID NO. 9 is the DNA sequence of TbRa19.

SEQ. ID NO. 10 is the DNA sequence of TbRa24.

SEQ. ID NO. 11 is the DNA sequence of TbRa26.

SEQ. ID NO. 12 is the DNA sequence of TbRa28.

SEQ. ID NO. 13 is the DNA sequence of TbRa29.

SEQ. ID NO. 14 is the DNA sequence of TbRa2A.

SEQ. ID NO. 15 is the DNA sequence of TbRa3.

SEQ. ID NO. 16 is the DNA sequence of TbRa32.

SEQ. ID NO. 17 is the DNA sequence of TbRa35.

SEQ. ID NO. 18 is the DNA sequence of TbRa36.

SEQ. ID NO. 19 is the DNA sequence of TbRa4.

SEQ. ID NO. 20 is the DNA sequence of TbRa9.

SEQ. ID NO. 21 is the DNA sequence of TbRaB.

SEQ. ID NO. 22 is the DNA sequence of TbRaC.

SEQ. ID NO. 23 is the DNA sequence of TbRaD

SEQ. ID NO. 24 is the DNA sequence of YYWCPG

SEQ. ID NO. 25 is the DNA sequence of AAMK.

SEQ. ID NO. 26 is the DNA sequence of TbL-23.

SEQ. ID NO. 27 is the DNA sequence of TbL-24.

SEQ. ID NO. 28 is the DNA sequence of TbL-25

SEQ. ID NO. 29 is the DNA sequence of TbL-28.

SEQ. ID NO. 35 is the DNA sequence of TbM-3.

SEQ. ID NO. 36 is the DNA sequence of TbM-6.

SEQ. ID NO. 37 is the DNA sequence of TbM-7.

SEQ. ID NO. 38 is the DNA sequence of TbM-9.

SEQ. ID NO. 39 is the DNA sequence of TbM-12.

SEQ. ID NO. 40 is the DNA sequence of TbM-13

SEQ. ID NO. 41 is the DNA sequence of TbM-14.

SEQ. ID NO. 42 is the DNA sequence of TbM-15.

SEQ. ID NO. 43 is the DNA sequence of TbH-4.

SEQ. ID NO. 44 is the DNA sequence of TbH-4-FWD.

SEQ. ID NO. 45 is the DNA sequence of TbH-12.

SEQ. ID NO. 46 is the DNA sequence of Tb38-1.

SEQ. ID NO. 47 is the DNA sequence of Tb38-4.

SEQ. ID NO. 48 is the DNA sequence of TbL-17.

SEQ. ID NO. 49 is the DNA sequence of TbL-20.

SEQ. ID NO. 50 is the DNA sequence of TbL-21.

SEQ. ID NO. 51 is the DNA sequence of TbH-16.

SEQ. ID NO. 52 is the DNA sequence of DPEP.

SEQ. ID NO. 53 is the deduced amino acid sequence of DPEP.

SEQ. ID NO. 54 is the protein sequence of DPV N-terminal Antigen.

SEQ. ID NO. 55 is the protein sequence of AVGS N-terminal Antigen.

SEQ. ID NO. 56 is the protein sequence of AAMK N-terminal Antigen.

SEQ. ID NO. 57 is the protein sequence of YYWC N-terminal Antigen.

SEQ. ID NO. 58 is the protein sequence of DIGS N-terminal Antigen.

SEQ. ID NO. 59 is the protein sequence of AFES N-terminal Antigen.

SEQ. ID NO. [illegible] is the [illegible] amino acid sequence of [illegible] peptide.

SEQ. ID NO. [illegible]

SEQ. ID NO. 65 is the deduced amino acid sequence of TbRa10.

SEQ. ID NO. 66 is the deduced amino acid sequence of TbRa11.

SEQ. ID NO. 67 is the deduced amino acid sequence of TbRa12.

SEQ. ID NO. 68 is the deduced amino acid sequence of TbRa13.

SEQ. ID NO. 69 is the deduced amino acid sequence of TbRa16.

SEQ. ID NO. 70 is the deduced amino acid sequence of TbRa17.

SEQ. ID NO. 71 is the deduced amino acid sequence of TbRa18.

SEQ. ID NO. 72 is the deduced amino acid sequence of TbRa19.

SEQ. ID NO. 73 is the deduced amino acid sequence of TbRa24.

SEQ. ID NO. 74 is the deduced amino acid sequence of TbRa26.

SEQ. ID NO. 75 is the deduced amino acid sequence of TbRa28.

SEQ. ID NO. 76 is the deduced amino acid sequence of TbRa29.

SEQ. ID NO. 77 is the deduced amino acid sequence of TbRa2A.

SEQ. ID NO. 78 is the deduced amino acid sequence of TbRa3.

SEQ. ID NO. 79 is the deduced amino acid sequence of TbRa32.

SEQ. ID NO. 80 is the deduced amino acid sequence of TbRa35.

SEQ. ID NO. 81 is the deduced amino acid sequence of TbRa36.

SEQ. ID NO. 82 is the deduced amino acid sequence of TbRa4.

SEQ. ID NO. 83 is the deduced amino acid sequence of TbRa9.

SEQ. ID NO. 84 is the deduced amino acid sequence of TbRaB.

SEQ. ID NO. 85 is the deduced amino acid sequence of TbRaC.

SEQ. ID NO. 86 is the deduced amino acid sequence of TbRaD.

SEQ. ID NO. 87 is the deduced amino acid sequence of YYWCPG.

SEQ. ID NO. 88 is the deduced amino acid sequence of TbAAMK.

SEQ. ID NO. 89 is the deduced amino acid sequence of Tb38-1

SEQ. ID NO. 95 is the deduced amino acid sequence of DPAS.

SEQ. ID NO. 96 is the DNA sequence of DPV.

SEQ. ID NO. 97 is the deduced amino acid sequence of DPV.

SEQ. ID NO. 98 is the DNA sequence of ESAT-6.

SEQ. ID NO. 99 is the deduced amino acid sequence of ESAT-6.

SEQ. ID NO. 100 is the DNA sequence of TbH-8-2.

SEQ. ID NO. 101 is the DNA sequence of TbH-9FL.

SEQ. ID NO. 102 is the deduced amino acid sequence of TbH-9FL.

SEQ. ID NO. 103 is the DNA sequence of TbH-9-1.

SEQ. ID NO. 104 is the deduced amino acid sequence of TbH-9-1.

SEQ. ID NO. 105 is the DNA sequence of TbH-9-4.

SEQ. ID NO. 106 is the deduced amino acid sequence of TbH-9-4.

SEQ. ID NO. 107 is the DNA sequence of Tb38-1F2 IN.

SEQ. ID NO. 108 is the DNA sequence of Tb38-1F2 RP.

SEQ. ID NO. 109 is the deduced amino acid sequence of Tb37-FL.

SEQ. ID NO. 110 is the deduced amino acid sequence of Tb38-IN.

SEQ. ID NO. 111 is the DNA sequence of Tb38-1F3.

SEQ. ID NO. 112 is the deduced amino acid sequence of Tb38-1F3.

SEQ. ID NO. 113 is the DNA sequence of Tb38-1F5.

SEQ. ID NO. 114 is the DNA sequence of Tb38-1F6.

SEQ. ID NO. 115 is the deduced N-terminal amino acid sequence of DPV.

SEQ. ID NO. 116 is the deduced N-terminal amino acid sequence of AVGS.

SEQ. ID NO. 117 is the deduced N-terminal amino acid sequence of AAMK.

SEQ. ID NO. 118 is the deduced N-terminal amino acid sequence of YYWC

SEQ. ID NO. 119 is the deduced N-terminal amino acid sequence of DIGS.

SEQ. ID NO. [illegible]

[illegible] is the deduced N-terminal amino acid sequence of DPAS

SEQ. ID NO. [illegible] is the deduced N-terminal amino acid sequence of DPPD

SEQ ID NO. 125-128 are the protein sequences of four DPPD cyanogen bromide fragments.

SEQ ID NO. 129 is the N-terminal protein sequence of XDS antigen.

SEQ ID NO. 130 is the N-terminal protein sequence of AGD antigen.

SEQ ID NO. 131 is the N-terminal protein sequence of APE antigen.

SEQ ID NO. 132 is the N-terminal protein sequence of XYI antigen.

SEQ ID NO. 133 is the DNA sequence of TbH-29.

SEQ ID NO. 134 is the DNA sequence of TbH-30.

SEQ ID NO. 135 is the DNA sequence of TbH-32.

SEQ ID NO. 136 is the DNA sequence of TbH-33.

SEQ ID NO. 137 is the predicted amino acid sequence of TbH-29.

SEQ ID NO. 138 is the predicted amino acid sequence of TbH-30.

SEQ ID NO. 139 is the predicted amino acid sequence of TbH-32.

SEQ ID NO. 140 is the predicted amino acid sequence of TbH-33.

SEQ ID NO: 141-146 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 147 is the DNA sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 148 is the amino acid sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 149 is the DNA sequence of the M. tuberculosis antigen 38 kD.

SEQ ID NO: 150 is the amino acid sequence of the M. tuberculosis antigen 38 kD.

SEQ ID NO: 151 is the DNA sequence of XP14.

SEQ ID NO: 152 is the DNA sequence of XP24.

SEQ ID NO: 153 is the DNA sequence of XP31.

SEQ ID NO: 158 is the DNA sequence of XP27.

SEQ ID NO: 159 is the DNA sequence of XP36.

SEQ ID NO: 160 is the 5' DNA sequence of XP4.

SEQ ID NO: 161 is the 5' DNA sequence of XP5.

SEQ ID NO: 162 is the 5' DNA sequence of XP17.

SEQ ID NO: 163 is the 5' DNA sequence of XP30

SEQ ID NO: 164 is the 5' DNA sequence of XP2.

SEQ ID NO: 165 is the 3' DNA sequence of XP2.

SEQ ID NO: 166 is the 5' DNA sequence of XP3.

SEQ ID NO: 167 is the 3' DNA sequence of XP3.

SEQ ID NO: 168 is the 5' DNA sequence of XP6.

SEQ ID NO: 169 is the 3' DNA sequence of XP6.

SEQ ID NO: 170 is the 5' DNA sequence of XP18.

SEQ ID NO: 171 is the 3' DNA sequence of XP18.

SEQ ID NO: 172 is the 5' DNA sequence of XP19.

SEQ ID NO: 173 is the 3' DNA sequence of XP19.

SEQ ID NO: 174 is the 5' DNA sequence of XP22.

SEQ ID NO: 175 is the 3' DNA sequence of XP22.

SEQ ID NO: 176 is the 5' DNA sequence of XP25

SEQ ID NO: 177 is the 3' DNA sequence of XP25.

SEQ ID NO: 178 is the full-length DNA sequence of TbH4-XP1.

SEQ ID NO: 179 is the predicted amino acid sequence of TbH4-XP1.

SEQ ID NO: 180 is the predicted amino acid sequence encoded by the reverse complement of TbH4-XP1.

SEQ ID NO: 181 is a first predicted amino acid sequence encoded by XP36.

SEQ ID NO: 182 is [illegible]

[illegible]

SEQ ID NO: [illegible]

SEQ ID NO: 186 is the DNA sequence of RDIF8.

SEQ ID NO: 187 is the DNA sequence of RDIF10.

SEQ ID NO: 188 is the DNA sequence of RDIF11.

SEQ ID NO: 189 is the predicted amino acid sequence of RDIF2.

SEQ ID NO: 190 is the predicted amino acid sequence of RDIF5.

SEQ ID NO: 191 is the predicted amino acid sequence of RDIF8.

SEQ ID NO: 192 is the predicted amino acid sequence of RDIF10.

SEQ ID NO: 193 is the predicted amino acid sequence of RDIF11.

SEQ ID NO: 194 is the 5' DNA sequence of RDIF12.

SEQ ID NO: 195 is the 3' DNA sequence of RDIF12.

SEQ ID NO: 196 is the DNA sequence of RDIF7.

SEQ ID NO: 197 is the predicted amino acid sequence of RDIF7.

SEQ ID NO: 198 is the DNA sequence of DIF2-1.

SEQ ID NO: 199 is the predicted amino acid sequence of DIF2-1.

SEQ ID NO: 200-207 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD, Tb38-1 and DPEP (hereinafter referred to as TbF-2).

SEQ ID NO: 208 is the DNA sequence of the fusion protein TbF-2.

SEQ ID NO: 209 is the amino acid sequence of the fusion protein TbF-2.

DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is generally directed to compositions and methods for diagnosing tuberculosis. The compositions of the subject invention include polypeptides that comprise at least one antigenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and or modifications

a polypeptide comprising an antigenic portion of one of the above antigens may consist entirely of the antigenic portion, or may contain additional sequences. The additional sequences may be derived from the native *M. tuberculosis* antigen or may be heterologous, and such sequences may (but need not) be antigenic.

An "antigenic portion" of an antigen (which may or may not be soluble) is a portion that is capable of reacting with sera obtained from an *M. tuberculosis*-infected individual (*i.e.*, generates an absorbance reading with sera from infected individuals that is at least three standard deviations above the absorbance obtained with sera from uninfected individuals, in a representative ELISA assay described herein). An "*M. tuberculosis*-infected individual" is a human who has been infected with *M. tuberculosis* (*e.g.*, has an intradermal skin test response to PPD that is at least 0.5 cm in diameter). Infected individuals may display symptoms of tuberculosis or may be free of disease symptoms. Polypeptides comprising at least an antigenic portion of one or more *M. tuberculosis* antigens as described herein may generally be used, alone or in combination, to detect tuberculosis in a patient.

The compositions and methods of this invention also encompass variants of the above polypeptides. A "variant," as used herein, is a polypeptide that differs from the native antigen only in conservative substitutions and/or modifications, such that the antigenic properties of the polypeptide are retained. Such variants may generally be identified by modifying one of the above polypeptide sequences, and evaluating the antigenic properties of the modified polypeptide using, for example, the representative procedures described herein.

A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val

translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (*e.g.*, poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

In a related aspect, combination polypeptides are disclosed. A "combination polypeptide" is a polypeptide comprising at least one of the above antigenic portions and one or more additional antigenic *M. tuberculosis* sequences, which are joined via a peptide linkage into a single amino acid chain. The sequences may be joined directly (*i.e.*, with no intervening amino acids) or may be joined by way of a linker sequence (*e.g.*, Gly-Cys-Gly) that does not significantly diminish the antigenic properties of the component polypeptides.

In general, *M. tuberculosis* antigens, and DNA sequences encoding such antigens, may be prepared using any of a variety of procedures. For example, soluble antigens may be isolated from *M. tuberculosis* culture filtrate by procedures known to those of ordinary skill in the art, including anion-exchange and reverse phase chromatography. Purified antigens may then be evaluated for a desired property, such as the ability to react with sera obtained from an *M. tuberculosis*-infected individual. Such screens may be performed using the representative methods described herein. Antigens may then be partially sequenced using, for example, traditional Edman chemistry. See Edman and Berg, *Eur. J. Biochem.* 80:116-132, 1967.

Antigens may also be produced recombinantly using a DNA sequence that encodes the antigen, which has been inserted into an expression vector and expressed in an appropriate host. DNA molecules encoding soluble antigens may be isolated by screening an appropriate *M. tuberculosis* expression library with anti-sera (*e.g.*, rabbit) raised specifically against soluble *M. tuberculosis* antigens. DNA sequences encoding antigens that may or may

DNA sequences encoding soluble antigens may also be obtained by screening an appropriate *M. tuberculosis* cDNA or genomic DNA library for DNA sequences that hybridize to degenerate oligonucleotides derived from partial amino acid sequences of isolated soluble antigens. Degenerate oligonucleotide sequences for use in such a screen may be designed and synthesized, and the screen may be performed, as described (for example) in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY (and references cited therein). Polymerase chain reaction (PCR) may also be employed, using the above oligonucleotides in methods well known in the art, to isolate a nucleic acid probe from a cDNA or genomic library. The library screen may then be performed using the isolated probe.

Regardless of the method of preparation, the antigens described herein are "antigenic." More specifically, the antigens have the ability to react with sera obtained from an *M. tuberculosis*-infected individual. Reactivity may be evaluated using, for example, the representative ELISA assays described herein, where an absorbance reading with sera from infected individuals that is at least three standard deviations above the absorbance obtained with sera from uninfected individuals is considered positive.

Antigenic portions of *M. tuberculosis* antigens may be prepared and identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3d ed., Raven Press, 1993, pp. 243-247 and references cited therein. Such techniques include screening polypeptide portions of the native antigen for antigenic properties. The representative ELISAs described herein may generally be employed in these screens. An antigenic portion of a polypeptide is a portion that, within such representative assays, generates a signal in such assays that is substantially similar to that generated by the full length antigen. In other words, an antigenic portion of a *M. tuberculosis* antigen generates at least about 20%, and preferably about 100%, of the signal induced by the full length antigen

commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. *See* Merrifield, *J. Am. Chem. Soc. 85*:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Applied BioSystems, Inc., Foster City, CA, and may be operated according to the manufacturer's instructions. Variants of a native antigen may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis. Sections of the DNA sequence may also be removed using standard techniques to permit preparation of truncated polypeptides.

Recombinant polypeptides containing portions and/or variants of a native antigen may be readily prepared from a DNA sequence encoding the polypeptide using a variety of techniques well known to those of ordinary skill in the art. For example, supernatants from suitable host/vector systems which secrete recombinant protein into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant protein.

Any of a variety of expression vectors known to those of ordinary skill in the art may be employed to express recombinant polypeptides as described herein. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line, such as COS or CHO. The DNA sequences expressed in this manner may encode naturally occurring antigens, portions of naturally occurring antigens, or other variants thereof.

In general, regardless of the method of preparation, the polypeptides disclosed herein are prepared in substantially pure form. Preferably, the

In certain specific embodiments, the subject invention discloses polypeptides comprising at least an antigenic portion of a soluble *M. tuberculosis* antigen (or a variant of such an antigen), where the antigen has one of the following N-terminal sequences:

(a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 115),

(b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);

(c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 117);

(d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);

(e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);

(f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 120);

(g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser (SEQ ID NO: 121);

(h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);

(i) Asp Pro Ala Ser Ala Pro Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-Thr-Ser-Leu Leu-Asn Ser-Leu-Ala Asp Pro Asn Val Ser Phe-Ala-Asn (SEQ ID NO: 123);

(j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID NO: 129)

(k) Ala-Gly-Asp-Thr-Xaa-Ile-Lys-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp,

amino acid sequence of which is provided in SEQ ID NO: 53. A DNA sequence encoding the antigen identified as (a) above is provided in SEQ ID NO: 96; its deduced amino acid sequence is provided in SEQ ID NO: 97. A DNA sequence corresponding to antigen (d) above is provided in SEQ ID NO: 24, a DNA sequence corresponding to antigen (c) is provided in SEQ ID NO: 25 and a DNA sequence corresponding to antigen (I) is disclosed in SEQ ID NO: 94 and its deduced amino acid sequence is provided in SEQ ID NO: 95.

In a further specific embodiment, the subject invention discloses polypeptides comprising at least an immunogenic portion of an *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only in conservative substitutions and/or modifications:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132) or

(n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124)

wherein Xaa may be any amino acid, preferably a cysteine residue.

In other specific embodiments, the subject invention discloses polypeptides comprising at least an antigenic portion of a soluble *M. tuberculosis* antigen (or a variant of such an antigen) that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID NOS: 1, 2, 4 10, 13 25, 52, 94 and 96, (b) the complements of such DNA sequences, or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

In further specific embodiments, the subject invention discloses polypeptides comprising at least an antigenic portion of a *M. tuberculosis* antigen (or a variant of such an antigen), which may or may not be soluble, that comprises one or more of the amino acid

In [illegible] embodiments discussed above, the *M. tuberculosis* antigens

or more of DNA sequences specifically recited herein. "Substantial homology," as used herein, refers to DNA sequences that are capable of hybridizing under moderately stringent conditions. Suitable moderately stringent conditions include prewashing in a solution of 5X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5X SSC, overnight or, in the event of cross-species homology, at 45°C with 0.5X SSC; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS). Such hybridizing DNA sequences are also within the scope of this invention, as are nucleotide sequences that, due to code degeneracy, encode an immunogenic polypeptide that is encoded by a hybridizing DNA sequence.

In a related aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, a polypeptide of the present invention and a known *M. tuberculosis* antigen, such as the 38 kD antigen described above or ESAT-6 (SEQ ID NOS: 98 and 99), together with variants of such fusion proteins. The fusion proteins of the present invention may also include a linker peptide between the first and second polypeptides.

A DNA sequence encoding a fusion protein of the present invention is constructed using known recombinant DNA techniques to assemble separate DNA sequences encoding the first and second polypeptides into an appropriate expression vector. The 3' end of a DNA sequence encoding the first polypeptide is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide so that the reading frames of the sequences are in phase to permit mRNA translation of the two DNA sequences into a single fusion protein that retains the biological activity of both the first and the second polypeptides.

A peptide linker sequence may be employed to separate the first and the second polypeptides by a distance sufficient to ensure that each polypeptide folds into its

or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene 40*:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA 83*:8258-8562, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may be from 1 to about 50 amino acids in length. Peptide linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric hindrance.

In another aspect, the present invention provides methods for using the polypeptides described above to diagnose tuberculosis. In this aspect, methods are provided for detecting *M. tuberculosis* infection in a biological sample, using one or more of the above polypeptides, alone or in combination. In embodiments in which multiple polypeptides are employed, polypeptides other than those specifically described herein, such as the 38 kD antigen described in Andersen and Hansen, *Infect. Immun. 57*:2481-2488, 1989, may be included. As used herein, a "biological sample" is any antibody-containing sample obtained from a patient. Preferably, the sample is whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid or urine. More preferably, the sample is a blood, serum or plasma sample obtained from a patient or a blood supply. The polypeptide(s) are used in an assay, as described below, to determine the presence or absence of antibodies to the polypeptide(s) in the sample, relative to a predetermined cut off value. The presence of such antibodies indicates previous sensitization to mycobacterial antigens which may be indicative of tuberculosis.

In embodiments in which more than one polypeptide is employed, the polypeptides used are preferably complementary (*i.e.*, one component polypeptide will tend

formulated that are capable of detecting infection in most, or all, of the samples tested. Such polypeptides are complementary. For example, approximately 25-30% of sera from tuberculosis-infected individuals are negative for antibodies to any single protein, such as the 38 kD antigen mentioned above. Complementary polypeptides may, therefore, be used in combination with the 38 kD antigen to improve sensitivity of a diagnostic test.

There are a variety of assay formats known to those of ordinary skill in the art for using one or more polypeptides to detect antibodies in a sample. *See*, *e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988, which is incorporated herein by reference. In a preferred embodiment, the assay involves the use of polypeptide immobilized on a solid support to bind to and remove the antibody from the sample. The bound antibody may then be detected using a detection reagent that contains a reporter group. Suitable detection reagents include antibodies that bind to the antibody/polypeptide complex and free polypeptide labeled with a reporter group (*e.g.*, in a semi-competitive assay). Alternatively, a competitive assay may be utilized, in which an antibody that binds to the polypeptide is labeled with a reporter group and allowed to bind to the immobilized antigen after incubation of the antigen with the sample. The extent to which components of the sample inhibit the binding of the labeled antibody to the polypeptide is indicative of the reactivity of the sample with the immobilized polypeptide.

The solid support may be any solid material known to those of ordinary skill in the art to which the antigen may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681.

The polypeptides may be bound to the solid support using a variety of

membrane is preferred. In such cases, adsorption may be achieved by contacting the polypeptide, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of polypeptide ranging from about 10 ng to about 1 μg, and preferably about 100 ng, is sufficient to bind an adequate amount of antigen.

Covalent attachment of polypeptide to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the polypeptide. For example, the polypeptide may be bound to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the polypeptide (*see, e.g.*, Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is an enzyme linked immunosorbent assay (ELISA). This assay may be performed by first contacting a polypeptide antigen that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that antibodies to the polypeptide within the sample are allowed to bind to the immobilized polypeptide. Unbound sample is then removed from the immobilized polypeptide and a detection reagent capable of binding to the immobilized antibody-polypeptide complex is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific detection reagent.

More specifically, once the polypeptide is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO) may be employed. The

[illegible] method [illegible] detect the presence of antibody [illegible]

of binding that is at least 95% of that achieved at equilibrium between bound and unbound antibody. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. Detection reagent may then be added to the solid support. An appropriate detection reagent is any compound that binds to the immobilized antibody-polypeptide complex and that can be detected by any of a variety of means known to those in the art. Preferably, the detection reagent contains a binding agent (such as, for example, Protein A, Protein G, immunoglobulin, lectin or free antigen) conjugated to a reporter group. Preferred reporter groups include enzymes (such as horseradish peroxidase), substrates, cofactors, inhibitors, dyes, radionuclides, luminescent groups, fluorescent groups and biotin. The conjugation of binding agent to reporter group may be achieved using standard methods known to those of ordinary skill in the art. Common binding agents may also be purchased conjugated to a variety of reporter groups from many commercial sources (*e.g.*, Zymed Laboratories, San Francisco, CA, and Pierce, Rockford, IL).

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound antibody. An appropriate amount of time may generally be determined from the manufacturer's instructions or by assaying the level of binding that occurs over a period of time. Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods

To determine the presence or absence of anti-*M. tuberculosis* antibodies in the sample, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value is the average mean signal obtained when the immobilized antigen is incubated with samples from an uninfected patient. In general, a sample generating a signal that is three standard deviations above the predetermined cut-off value is considered positive for tuberculosis. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology. A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, pp. 106-107. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for tuberculosis.

In a related embodiment, the assay is performed in a rapid flow-through or strip test format, wherein the antigen is immobilized on a membrane, such as nitrocellulose. In the flow-through test, antibodies within the sample bind to the immobilized polypeptide as the sample passes through the membrane. A detection reagent (*e.g.*, protein A-colloidal gold) then binds to the antibody polypeptide complex as the solution containing the detection reagent flows through the membrane. The detection of bound detection reagent may then be

[illegible] of detection reagent at the polypeptide indicates the

detection reagent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of polypeptide immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of antibodies that would be sufficient to generate a positive signal in an ELISA, as discussed above. Preferably, the amount of polypeptide immobilized on the membrane ranges from about 25 ng to about 1 μg, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount (*e.g.*, one drop) of patient serum or blood.

Of course, numerous other assay protocols exist that are suitable for use with the polypeptides of the present invention. The above descriptions are intended to be exemplary only.

In yet another aspect, the present invention provides antibodies to the inventive polypeptides. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. *See, e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one such technique, an immunogen comprising the antigenic polypeptide is initially injected into any of a wide variety of mammals (*e.g.*, mice, rats, rabbits, sheep and goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

Antibodies may be used in diagnostic tests to detect the presence of *M tuberculosis* antigens using assays similar to those detailed above and other techniques well known to those of skill in the art, thereby providing a method for detecting *M tuberculosis* infection in a patient.

Diagnostic reagents of the present invention may also comprise DNA sequences encoding one or more of the above polypeptides, or one or more portions thereof. For example, at least two oligonucleotide primers may be employed in a polymerase chain

[illegible] cDNA [illegible] as is well known in the art, such as by electrophoresis [illegible]

present invention may be used in a hybridization assay to detect the presence of an inventive polypeptide in a biological sample.

As used herein, the term "oligonucleotide primer/probe specific for a DNA molecule" means an oligonucleotide sequence that has at least about 80%, preferably at least about 90% and more preferably at least about 95%, identity to the DNA molecule in question. Oligonucleotide primers and/or probes which may be usefully employed in the inventive diagnostic methods preferably have at least about 10-40 nucleotides. In a preferred embodiment, the oligonucleotide primers comprise at least about 10 contiguous nucleotides of a DNA molecule encoding one of the polypeptides disclosed herein. Preferably, oligonucleotide probes for use in the inventive diagnostic methods comprise at least about 15 contiguous oligonucleotides of a DNA molecule encoding one of the polypeptides disclosed herein. Techniques for both PCR based assays and hybridization assays are well known in the art (see, for example, Mullis *et al. Ibid*; Ehrlich, *Ibid*). Primers or probes may thus be used to detect *M. tuberculosis*-specific sequences in biological samples. DNA probes or primers comprising oligonucleotide sequences described above may be used alone, in combination with each other, or with previously identified sequences, such as the 38 kD antigen discussed above.

The following Examples are offered by way of illustration and not by way of limitation

EXAMPLES

EXAMPLE 1

PURIFICATION AND CHARACTERIZATION OF POLYPEPTIDES

[illegible]

M. tuberculosis (either H37Ra, ATCC No. 25177, or H37Rv, ATCC No. 25618) was cultured in sterile GAS media at 37°C for fourteen days. The media was then vacuum filtered (leaving the bulk of the cells) through a 0.45 μ filter into a sterile 2.5 L bottle. The media was then filtered through a 0.2 μ filter into a sterile 4 L bottle. NaN_3 was then added to the culture filtrate to a concentration of 0.04%. The bottles were then placed in a 4°C cold room.

The culture filtrate was concentrated by placing the filtrate in a 12 L reservoir that had been autoclaved and feeding the filtrate into a 400 ml Amicon stir cell which had been rinsed with ethanol and contained a 10,000 kDa MWCO membrane. The pressure was maintained at 60 psi using nitrogen gas. This procedure reduced the 12 L volume to approximately 50 ml.

The culture filtrate was then dialyzed into 0.1% ammonium bicarbonate using a 8,000 kDa MWCO cellulose ester membrane, with two changes of ammonium bicarbonate solution. Protein concentration was then determined by a commercially available BCA assay (Pierce, Rockford, IL).

The dialyzed culture filtrate was then lyophilized, and the polypeptides resuspended in distilled water. The polypeptides were then dialyzed against 0.01 mM 1,3 bis[tris(hydroxymethyl)-methylamino]propane, pH 7.5 (Bis-Tris propane buffer), the initial conditions for anion exchange chromatography. Fractionation was performed using gel profusion chromatography on a POROS 146 II Q/M anion exchange column 4.6 mm x 100 mm (Perseptive BioSystems, Framingham, MA) equilibrated in 0.01 mM Bis-Tris propane buffer pH 7.5. Polypeptides were eluted with a linear 0-0.5 M NaCl gradient in the above buffer system. The column eluent was monitored at a wavelength of 220 nm.

The pools of polypeptides eluting from the ion exchange column were dialyzed against distilled water and lyophilized. The resulting material was dissolved in 0.1% trifluoroacetic acid (TFA) and fractionated by reverse phase HPLC using a Vydac

to maximize the purity of the individual samples. Approximately 200 purified polypeptides were obtained.

The purified polypeptides were then screened for the ability to induce T-cell proliferation in PBMC preparations. The PBMCs from donors known to be PPD skin test positive and whose T cells were shown to proliferate in response to PPD and crude soluble proteins from MTB were cultured in medium comprising RPMI 1640 supplemented with 10% pooled human serum and 50 μg/ml gentamicin. Purified polypeptides were added in duplicate at concentrations of 0.5 to 10 μg/ml. After six days of culture in 96-well round-bottom plates in a volume of 200 μl, 50 μl of medium was removed from each well for determination of IFN-γ levels, as described below. The plates were then pulsed with 1 μCi/well of tritiated thymidine for a further 18 hours, harvested and tritium uptake determined using a gas scintillation counter. Fractions that resulted in proliferation in both replicates three fold greater than the proliferation observed in cells cultured in medium alone were considered positive.

IFN-γ was measured using an enzyme-linked immunosorbent assay (ELISA). ELISA plates were coated with a mouse monoclonal antibody directed to human IFN-γ (Chemicon) in PBS for four hours at room temperature. Wells were then blocked with PBS containing 5% (W/V) non-fat dried milk for 1 hour at room temperature. The plates were then washed six times in PBS/0.2% TWEEN-20 and samples diluted 1:2 in culture medium in the ELISA plates were incubated overnight at room temperature. The plates were again washed and a polyclonal rabbit anti-human IFN-γ serum diluted 1:3000 in PBS/10% normal goat serum was added to each well. The plates were then incubated for two hours at room temperature, washed and horseradish peroxidase-coupled anti-rabbit IgG (Jackson Labs.) was added at a 1:2000 dilution in PBS/5% non-fat dried milk. After a further two hour incubation at room temperature, the plates were washed and TMB substrate added. The reaction was

For sequencing, the polypeptides were individually dried onto Biobrene™ (Perkin Elmer/Applied BioSystems Division, Foster City, CA) treated glass fiber filters. The filters with polypeptide were loaded onto a Perkin Elmer/Applied BioSystems Division Procise 492 protein sequencer. The polypeptides were sequenced from the amino terminal and using traditional Edman chemistry. The amino acid sequence was determined for each polypeptide by comparing the retention time of the PTH amino acid derivative to the appropriate PTH derivative standards.

Using the procedure described above, antigens having the following N-terminal sequences were isolated:

(a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Xaa-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 54);

(b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 55);

(c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 56);

(d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 57);

(e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 58);

(f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 59);

(g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ala-Pro-Pro-Ala (SEQ ID NO: 60); and

(h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 61);

[illegible] purification step previously described was [illegible]

City, CA) with a 7 micron pore size, column size 1 mm x 100 mm, in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted from the column with a linear gradient of 1%/minute of acetonitrile (containing 0.05% TFA) in water (0.05% TFA) at a flow rate of 80 µl/minute. The eluent was monitored at 250 nm. The original fraction was separated into 4 major peaks plus other smaller components and a polypeptide was obtained which was shown to have a molecular weight of 12,054 Kd (by mass spectrometry) and the following N-terminal sequence:

(i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-Thr-Ser-Leu-Leu-Asn-Asn-Leu-Ala-Asp-Pro-Asp-Val-Ser-Phe-Ala-Asp (SEQ ID NO: 62).

This polypeptide was shown to induce proliferation and IFN-γ production in PBMC preparations using the assays described above.

Additional soluble antigens were isolated from *M. tuberculosis* culture filtrate as follows. *M. tuberculosis* culture filtrate was prepared as described above. Following dialysis against Bis-Tris propane buffer, at pH 5.5, fractionation was performed using anion exchange chromatography on a Poros QE column 4.6 x 100 mm (Perseptive Biosystems) equilibrated in Bis-Tris propane buffer pH 5.5. Polypeptides were eluted with a linear 0-1.5 M NaCl gradient in the above buffer system at a flow rate of 10 ml/min. The column eluent was monitored at a wavelength of 214 nm.

The fractions eluting from the ion exchange column were pooled and subjected to reverse phase chromatography using a Poros R2 column 4.6 x 100 mm (Perseptive Biosystems). Polypeptides were eluted from the column with a linear gradient from 0-100% acetonitrile (0.1% TFA) at a flow rate of 5 ml/min. The eluent was monitored at 214 nm.

Fractions containing the eluted polypeptides were lyophilized and resuspended

The fraction with biological activity was separated into one major peak plus other smaller components. Western blot of this peak onto PVDF membrane revealed three major bands of molecular weights 14 Kd, 20 Kd and 26 Kd. These polypeptides were determined to have the following N-terminal sequences, respectively:

(j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID NO: 129)

(k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID NO: 130) and

(l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID NO: 131), wherein Xaa may be any amino acid.

Using the assays described above, these polypeptides were shown to induce proliferation and IFN-γ production in PBMC preparations. Figs. 1A and B show the results of such assays using PBMC preparations from a first and a second donor, respectively.

DNA sequences that encode the antigens designated as (a)..(c), (d) and (g) above were obtained by screening a *M. tuberculosis* genomic library using ^{32}P end labeled degenerate oligonucleotides corresponding to the N-terminal sequence and containing *M. tuberculosis* codon bias. The screen performed using a probe corresponding to antigen (a) above identified a clone having the sequence provided in SEQ ID NO: 96. The polypeptide encoded by SEQ ID NO: 96 is provided in SEQ ID NO: 97. The screen performed using a probe corresponding to antigen (g) above identified a clone having the sequence provided in SEQ ID NO: 52. The polypeptide encoded by SEQ ID NO: 52 is provided in SEQ ID NO: 53. The screen performed using a probe corresponding to antigen (d) above identified a clone having the sequence provided in SEQ ID NO: 24, and the screen performed with a probe corresponding to antigen (e) identified a clone having the sequence provided in SEQ ID NO: 25

The amino acid sequence for antigen (i) was found to be homologous to a sequence from *M. leprae*. The full length *M. leprae* sequence was amplified from genomic DNA using the sequence obtained from GENBANK. This sequence was then used to screen an *M. tuberculosis* library and a full length copy of the *M. tuberculosis* homologue was obtained (SEQ ID NO: 94).

The amino acid sequence for antigen (j) was found to be homologous to a known *M. tuberculosis* protein translated from a DNA sequence. To the best of the inventors' knowledge, this protein has not been previously shown to possess T-cell stimulatory activity. The amino acid sequence for antigen (k) was found to be related to a sequence from *M. leprae*.

In the proliferation and IFN-γ assays described above, using three PPD positive donors, the results for representative antigens provided above are presented in Table 1:

TABLE 1

RESULTS OF PBMC PROLIFERATION AND IFN-γ ASSAYS

Sequence	Proliferation	IFN-γ
(a)	+	-
(c)	+++	+++
(d)	++	++
(g)	+++	+++
(h)	+++	+++

In Table 1, responses that gave a stimulation index (SI) of between 2 and 4

indicate that these antigens are capable of inducing proliferation and/or interferon-γ production.

EXAMPLE 2

USE OF PATIENT SERA TO ISOLATE *M. TUBERCULOSIS* ANTIGENS

This example illustrates the isolation of antigens from *M. tuberculosis* lysate by screening with serum from *M. tuberculosis*-infected individuals.

Dessicated *M. tuberculosis* H37Ra (Difco Laboratories) was added to a 2% NP40 solution, and alternately homogenized and sonicated three times. The resulting suspension was centrifuged at 13,000 rpm in microfuge tubes and the supernatant put through a 0.2 micron syringe filter. The filtrate was bound to Macro Prep DEAE beads (BioRad, Hercules, CA). The beads were extensively washed with 20 mM Tris pH 7.5 and bound proteins eluted with 1M NaCl. The NaCl elute was dialyzed overnight against 10 mM Tris, pH 7.5. Dialyzed solution was treated with DNase and RNase at 0.05 mg/ml for 30 min. at room temperature and then with α-D-mannosidase, 0.5 U/mg at pH 4.5 for 3-4 hours at room temperature. After returning to pH 7.5, the material was fractionated via FPLC over a Bio Scale-Q-20 column (BioRad). Fractions were combined into nine pools, concentrated in a Centriprep 10 (Amicon, Beverley, MA) and screened by Western blot for serological activity using a serum pool from *M. tuberculosis*-infected patients which was not immunoreactive with other antigens of the present invention.

The most reactive fraction was run in SDS-PAGE and transferred to PVDF. A band at approximately 85 Kd was cut out yielding the sequence:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn Val-His-Leu-Val; (SEQ ID NO: 132), wherein Xaa may be any

degenerate oligonucleotides corresponding to the N-terminal sequence of SEQ ID NO:137. A clone was identified having the DNA sequence provided in SEQ ID NO: 198. This sequence was found to encode the amino acid sequence provided in SEQ ID NO: 199. Comparison of these sequences with those in the genebank revealed some similarity to sequences previously identified in *M. tuberculosis* and *M. bovis*.

EXAMPLE 3

PREPARATION OF DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

This example illustrates the preparation of DNA sequences encoding *M. tuberculosis* antigens by screening a *M. tuberculosis* expression library with sera obtained from patients infected with *M. tuberculosis*, or with anti-sera raised against *M. tuberculosis* antigens.

A. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTI-SERA RAISED AGAINST M. TUBERCULOSIS SUPERNATANT

Genomic DNA was isolated from the *M. tuberculosis* strain H37Ra. The DNA was randomly sheared and used to construct an expression library using the Lambda ZAP expression system (Stratagene, La Jolla, CA). Rabbit anti-sera was generated against secretory proteins of the *M. tuberculosis* strains H37Ra, H37Rv and Erdman by immunizing a rabbit with concentrated supernatant of the *M. tuberculosis* cultures. Specifically, the rabbit was first immunized subcutaneously with 200 µg of protein antigen in a total volume of 2 ml containing 100 µg muramyl dipeptide (Calbiochem, La Jolla, CA) and 1 ml of incomplete Freund's adjuvant. Four weeks later the rabbit was boosted subcutaneously with 100 µg antigen in incomplete Freund's adjuvant. Finally, the rabbit was immunized intravenously

Thirty two clones were purified. Of these, 25 represent sequences that have not been previously identified in *M. tuberculosis*. Proteins were induced by IPTG and purified by gel elution, as described in Skeiky et al., *J. Exp. Med. 181*:1527-1537, 1995. Representative partial sequences of DNA molecules identified in this screen are provided in SEQ ID NOS: 1-25. The corresponding predicted amino acid sequences are shown in SEQ ID NOS: 64-88.

On comparison of these sequences with known sequences in the gene bank using the databases described above, it was found that the clones referred to hereinafter as TbRA2A, TbRA16, TbRA18, and TbRA29 (SEQ ID NOS: 77, 69, 71, 76) show some homology to sequences previously identified in *Mycobacterium leprae* but not in *M. tuberculosis*. TbRA11, TbRA26, TbRA28 and TbDPEP (SEQ ID NOS: 66, 74, 75, 53) have been previously identified in *M. tuberculosis*. No significant homologies were found to TbRA1, TbRA3, TbRA4, TbRA9, TbRA10, TbRA13, TbRA17, TbRA19, TbRA29, TbRA32, TbRA36 and the overlapping clones TbRA35 and TbRA12 (SEQ ID NOS: 64, 78, 82, 83, 65, 68, 76, 72, 76, 79, 81, 80, 67, respectively). The clone TbRa24 is overlapping with clone TbRa29.

B. USE OF SERA FROM PATIENTS HAVING PULMONARY OR PLEURAL TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

The genomic DNA library described above, and an additional H37Rv library, were screened using pools of sera obtained from patients with active tuberculosis. To prepare the H37Rv library, *M. tuberculosis* strain H37Rv genomic DNA was isolated, subjected to partial Sau3A digestion and used to construct an expression library using the Lambda Zap expression system (Stratagene, La Jolla, Ca). Three different pools of sera, each containing sera obtained from three individuals with active pulmonary or pleural disease, were used in

lacked increased reactivity with the recombinant 38 kD *M. tuberculosis* H37Ra phosphate-binding protein.

All pools were pre-adsorbed with *E. coli* lysate and used to screen the H37Ra and H37Rv expression libraries, as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

Thirty two clones were purified. Of these, 31 represented sequences that had not been previously identified in human *M. tuberculosis*. Representative sequences of the DNA molecules identified are provided in SEQ ID NOS:: 26-51 and 100. Of these, TbH-8-2 (SEQ. ID NO. 100) is a partial clone of TbH-8, and TbH-4 (SEQ. ID NO. 43) and TbH-4-FWD (SEQ. ID NO. 44) are non-contiguous sequences from the same clone. Amino acid sequences for the antigens hereinafter identified as Tb38-1, TbH-4, TbH-8, TbH-9, and TbH-12 are shown in SEQ ID NOS.: 89-93. Comparison of these sequences with known sequences in the gene bank using the databases identified above revealed no significant homologies to TbH-4, TbH-8, TbH-9 and TbM-3, although weak homologies were found to TbH-9. TbH-12 was found to be homologous to a 34 kD antigenic protein previously identified in *M. paratuberculosis* (Acc. No. S28515). Tb38-1 was found to be located 34 base pairs upstream of the open reading frame for the antigen ESAT-6 previously identified in *M. bovis* (Acc. No. U34848) and in *M. tuberculosis* (Sorensen et al., *Infec. Immun* 63:1710-1717, 1995).

Probes derived from Tb38-1 and TbH-9, both isolated from an H37Ra library, were used to identify clones in an H37Rv library. Tb38-1 hybridized to Tb38-1F2, Tb38-1F3, Tb38-1F5 and Tb38-1F6 (SEQ. ID NOS: 107, 108, 111, 113, and 114). (SEQ ID NOS: 107 and 108 are non-contiguous sequences from clone Tb38-1F2.) Two open reading frames

[illegible] H37Rv library. TbH-9-F1 (SEQ. ID NO. [illegible], which

ID NO. 105) is a partial clone of TbH-8. The deduced amino acid sequences for these three clones are presented in SEQ ID NOS: 102, 104 and 106.

Further screening of the *M. tuberculosis* genomic DNA library, as described above, resulted in the recovery of ten additional reactive clones, representing seven different genes. One of these genes was identified as the 38 Kd antigen discussed above, one was determined to be identical to the 14Kd alpha crystallin heat shock protein previously shown to be present in *M. tuberculosis*, and a third was determined to be identical to the antigen TbH-8 described above. The determined DNA sequences for the remaining five clones (hereinafter referred to as TbH-29, TbH-30, TbH-32 and TbH-33) are provided in SEQ ID NO: 133-136, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 137-140, respectively. The DNA and amino acid sequences for these antigens were compared with those in the gene bank as described above. No homologies were found to the 5' end of TbH-29 (which contains the reactive open reading frame), although the 3' end of TbH-29 was found to be identical to the *M. tuberculosis* cosmid Y227. TbH-32 and TbH-33 were found to be identical to the previously identified *M. tuberculosis* insertion element IS6110 and to the *M. tuberculosis* cosmid Y50, respectively. No significant homologies to TbH-30 were found.

Positive phagemid from this additional screening were used to infect *E. coli* XL-1 Blue MRF', as described in Sambrook et al., *supra*. Induction of recombinant protein was accomplished by the addition of IPTG. Induced and uninduced lysates were run in duplicate on SDS-PAGE and transferred to nitrocellulose filters. Filters were reacted with human *M. tuberculosis* sera (1:200 dilution) reactive with TbH and a rabbit sera (1:200 or 1:250 dilution) reactive with the N-terminal 4 Kd portion of lacZ. Sera incubations were performed for 2 hours at room temperature. Bound antibody was detected by addition of ^{125}I-labeled Protein A and subsequent exposure to film for variable times ranging from 16 hours

TABLE 2

Antigen	Human M. tb Sera	Anti-lacZ Sera
TbH-29	45 Kd	45 Kd
TbH 30	No reactivity	29 Kd
TbH-32	12 Kd	12 Kd
TbH-33	16 Kd	16 Kd

Positive reaction of the recombinant human *M. tuberculosis* antigens with both the human *M. tuberculosis* sera and anti-lacZ sera indicate that reactivity of the human *M. tuberculosis* sera is directed towards the fusion protein. Antigens reactive with the anti-lacZ sera but not with the human *M. tuberculosis* sera may be the result of the human *M. tuberculosis* sera recognizing conformational epitopes, or the antigen-antibody binding kinetics may be such that the 2 hour sera exposure in the immunoblot is not sufficient.

Studies were undertaken to determine whether the antigens TbH-9 and Tb38-1 represent cellular proteins or are secreted into *M. tuberculosis* culture media. In the first study, rabbit sera were raised against A) secretory proteins of *M. tuberculosis*, B) the known secretory recombinant *M. tuberculosis* antigen 85b, C) recombinant Tb38-1 and D) recombinant TbH-9, using protocols substantially as described in Example 3A. Total *M. tuberculosis* lysate, concentrated supernatant of *M. tuberculosis* cultures and the recombinant antigens 85b, TbH-9 and Tb38-1 were resolved on denaturing gels, immobilized on nitrocellulose membranes and duplicate blots were probed using the rabbit sera described above.

The results of this analysis using control sera raised [illegible]

residues and would therefore be expected to migrate with a mobility approximately 1 kD larger that the native protein. In Figure 2D, recombinant TbH-9 is lacking approximately 10 kD of the full-length 42 kD antigen, hence the significant difference in the size of the immunoreactive native TbH-9 antigen in the lysate lane (indicated by an arrow). These results demonstrate that Tb38-1 and TbH-9 are intracellular antigens and are not actively secreted by *M. tuberculosis*.

The finding that TbH-9 is an intracellular antigen was confirmed by determining the reactivity of TbH-9-specific human T cell clones to recombinant TbH-9, secretory *M. tuberculosis* proteins and PPD. A TbH-9-specific T cell clone (designated 131TbH-9) was generated from PBMC of a healthy PPD-positive donor. The proliferative response of 131TbH-9 to secretory proteins, recombinant TbH-9 and a control *M. tuberculosis* antigen, TbRa11, was determined by measuring uptake of tritiated thymidine, as described in Example 1. As shown in Figure 3A, the clone 131TbH-9 responds specifically to TbH-9, showing that TbH-9 is not a significant component of *M. tuberculosis* secretory proteins. Figure 3B shows the production of IFN-γ by a second TbH-9-specific T cell clone (designated PPD 800-10) prepared from PBMC from a healthy PPD-positive donor, following stimulation of the T cell clone with secretory proteins, PPD or recombinant TbH-9. These results further confirm that TbH-9 is not secreted by *M. tuberculosis*.

C____ USE OF SERA FROM PATIENTS HAVING EXTRAPULMONARY TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). The resulting library was screened using pools of sera obtained from individuals with extrapulmonary tuberculosis [illegible]

153, respectively, with the 5' and 3' DNA sequences for XP32 being provided in SEQ ID NOS: 154 and 155, respectively. The predicted amino acid sequence for XP14 is provided in SEQ ID NO: 156. The reverse complement of XP14 was found to encode the amino acid sequence provided in SEQ ID NO: 157.

Comparison of the sequences for the remaining 14 clones (hereinafter referred to as XP1-XP6, XP17-XP19, XP22, XP25, XP27, XP30 and XP36) with those in the genebank as described above, revealed no homologies with the exception of the 3' ends of XP2 and XP6 which were found to bear some homology to known *M. tuberculosis* cosmids. The DNA sequences for XP27 and XP36 are shown in SEQ ID NOS: 158 and 159, respectively, with the 5' sequences for XP4, XP5, XP17 and XP30 being shown in SEQ ID NOS: 160-163, respectively, and the 5' and 3' sequences for XP2, XP3, XP6, XP18, XP19, XP22 and XP25 being shown in SEQ ID NOS: 164 and 165; 166 and 167, 168 and 169; 170 and 171; 172 and 173; 174 and 175; and 176 and 177, respectively. XP1 was found to overlap with the DNA sequences for TbH4, disclosed above. The full-length DNA sequence for TbH4-XP1 is provided in SEQ ID NO: 178. This DNA sequence was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 179. The reverse complement of TbH4-XP1 was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 180. The DNA sequence for XP36 was found to contain two open reading frames encoding the amino acid sequence shown in SEQ ID NOS: 181 and 182, with the reverse complement containing an open reading frame encoding the amino acid sequence shown in SEQ ID NO: 183.

Recombinant XP1 protein was prepared as described above in Example 3B, with a metal ion affinity chromatography column being employed for purification. Recombinant XP1 was found to stimulate cell proliferation and IFN-γ production in T cells isolated from an *M. tuberculosis*-immune donors.

serological activity with a serum pool from *M. tuberculosis*-infected patients which showed little or no immunoreactivity with other antigens of the present invention. Rabbit anti-sera was generated against the most reactive fraction using the method described in Example 3A . The anti-sera was used to screen an *M. tuberculosis* Erdman strain genomic DNA expression library prepared as described above. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones determined.

Ten different clones were purified. Of these, one was found to be TbRa35, described above, and one was found to be the previously identified *M. tuberculosis* antigen, HSP60. Of the remaining eight clones, six (hereinafter referred to as RDIF2, RDIF5, RDIF8, RDIF10, RDIF11 and RDIF12) were found to bear some similarity to previously identified *M. tuberculosis* sequences. The determined DNA sequences for RDIF2, RDIF5, RDIF8, RDIF10 and RDIF11 are provided in SEQ ID NOS: 184-188, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NOS: 189-193, respectively. The 5' and 3' DNA sequences for RDIF12 are provided in SEQ ID NOS: 194 and 195, respectively. No significant homologies were found to the antigen RDIF-7. The determined DNA and predicted amino acid sequences for RDIF7 are provided in SEQ ID NOS: 196 and 197, respectively. One additional clone, referred to as RDIF6 was isolated, however, this was found to be identical to RDIF5.

Recombinant RDIF6, RDIF8, RDIF10 and RDIF11 were prepared as described above. These antigens were found to stimulate cell proliferation and IFN-γ production in T cells isolated from *M. tuberculosis*-immune donors.

EXAMPLE 4

PPD was prepared as published with some modification (Seibert, F. et al., Tuberculin purified protein derivative. Preparation and analyses of a large quantity for standard. The American Review of Tuberculosis 44:9-25, 1941). *M. tuberculosis* Rv strain was grown for 6 weeks in synthetic medium in roller bottles at 37°C. Bottles containing the bacterial growth were then heated to 100°C in water vapor for 3 hours. Cultures were sterile filtered using a 0.22 μ filter and the liquid phase was concentrated 20 times using a 3 kD cut-off membrane. Proteins were precipitated once with 50% ammonium sulfate solution and eight times with 25% ammonium sulfate solution. The resulting proteins (PPD) were fractionated by reverse phase liquid chromatography (RP-HPLC) using a C18 column (7.8 x 300 mM; Waters, Milford, MA) in a Biocad HPLC system (Perseptive Biosystems, Framingham, MA). Fractions were eluted from the column with a linear gradient from 0-100% buffer (0.1% TFA in acetonitrile). The flow rate was 10 ml/minute and eluent was monitored at 214 nm and 280 nm.

Six fractions were collected, dried, suspended in PBS and tested individually in *M. tuberculosis*-infected guinea pigs for induction of delayed type hypersensitivity (DTH) reaction. One fraction was found to induce a strong DTH reaction and was subsequently fractionated further by RP-HPLC on a microbore Vydac C18 column (Cat. No. 218TP5115) in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted with a linear gradient from 5-100% buffer (0.05% TFA in acetonitrile) with a flow rate of 80 μl/minute. Eluent was monitored at 215 nm. Eight fractions were collected and tested for induction of DTH in *M. tuberculosis*-infected guinea pigs. One fraction was found to induce strong DTH of about 16 mm induration. The other fractions did not induce detectable DTH. The positive fraction was submitted to SDS-PAGE gel electrophoresis and found to contain a single protein band of approximately 12 kD molecular weight.

This polypeptide, herein after referred to as DPPD, was sequenced from the

[illegible] four cyanogen bromide fragments of DPPD were

EXAMPLE 5

SYNTHESIS OF SYNTHETIC POLYPEPTIDES

Polypeptides may be synthesized on a Millipore 9050 peptide synthesizer using FMOC chemistry with HPTU (O-Benzotriazole N,N,N',N'-tetramethyluronium hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray mass spectrometry and by amino acid analysis.

This procedure was used to synthesize a TbM-1 peptide that contains one and a half repeats of a TbM-1 sequence. The TbM-1 peptide has the sequence GCGDRSGGNLDQIRLRRDRSGGNL (SEQ ID NO: 63).

EXAMPLE 6

USE OF REPRESENTATIVE ANTIGENS FOR SERODIAGNOSIS OF TUBERCULOSIS

This Example illustrates the diagnostic properties of several representative

[illegible] The pure collected [illegible] removed and the wells were blocked for [illegible]

times with PBS/0.1% Tween 20™. 50 µL sera, diluted 1:100 in PBS/0.1% Tween 20™/0.1% BSA, was then added to each well and incubated for 30 minutes at room temperature. The plates were then washed again five times with PBS/0.1% Tween 20™.

The enzyme conjugate (horseradish peroxidase - Protein A, Zymed, San Francisco, CA) was then diluted 1:10,000 in PBS/0.1% Tween 20™/0.1% BSA, and 50 µL of the diluted conjugate was added to each well and incubated for 30 minutes at room temperature. Following incubation, the wells were washed five times with PBS/0.1% Tween 20™. 100 µL of tetramethylbenzidine peroxidase (TMB) substrate (Kirkegaard and Perry Laboratories, Gaithersburg, MD) was added, undiluted, and incubated for about 15 minutes. The reaction was stopped with the addition of 100 µL of 1 N H_2SO_4 to each well, and the plates were read at 450 nm.

Figure 4 shows the ELISA reactivity of two recombinant antigens isolated using method A in Example 3 (TbRa3 and TbRa9) with sera from *M. tuberculosis* positive and negative patients. The reactivity of these antigens is compared to that of bacterial lysate isolated from *M tuberculosis* strain H37Ra (Difco, Detroit, MI). In both cases, the recombinant antigens differentiated positive from negative sera. Based on cut-off values obtained from receiver-operator curves, TbRa3 detected 56 out of 87 positive sera, and TbRa9 detected 111 out of 165 positive sera.

Figure 5 illustrates the ELISA reactivity of representative antigens isolated using method B of Example 3. The reactivity of the recombinant antigens TbH4, TbH12, Tb38-1 and the peptide TbM-1 (as described in Example 4) is compared to that of the 38 kD antigen described by Andersen and Hansen, *Infect. Immun.* 57:2481-2488, 1989. Again, all of the polypeptides tested differentiated positive from negative sera. Based on cut-off values obtained from receiver-operator curves, TbH4 detected 67 out of 126 positive sera, TbH12 detected 50 out of 125 positive sera, 38-1 detected 61 out of 101 positive sera and the TbM-1

to the reactivity of *M. tuberculosis* lysate and the 38 kD antigen. The results are presented in Table 3, below:

TABLE 3

REACTIVITY OF ANTIGENS WITH SERA FROM *M. TUBERCULOSIS* PATIENTS

Patient	Acid Fast Sputum	ELISA Values					
		Lysate	38kD	TbRa9	TbH12	TbH4	TbRa3
Tb01B93I-2	++++	1.853	0.634	0.998	1.022	1.030	1.314
Tb01B93I-19	++++	2.657	2.322	0.608	0.837	1.857	2.335
Tb01B93I-8	+++	2.703	0.527	0.492	0.281	0.501	2.002
Tb01B93I-10	+++	1.665	1.301	0.685	0.216	0.448	0.458
Tb01B93I-11	+++	2.817	0.697	0.509	0.301	0.173	2.608
Tb01B93I-15	+++	1.28	0.283	0.808	0.218	1.537	0.811
Tb01B93I-16	+++	2.908	>3	0.899	0.441	0.593	1.080
Tb01B93I-25	+++	0.395	0.131	0.335	0.211	0.107	0.948
Tb01B93I-87	+++	2.653	2.432	2.282	0.977	1.221	0.857
Tb01B93I-89	++	1.912	2.370	2.436	0.876	0.520	0.952
Tb01B94I-108	+++	1.639	0.341	0.797	0.368	0.654	0.798
Tb01B94I-201	+++	1.721	0.419	0.661	0.137	0.064	0.692
Tb01B93I-88	++	1.939	1.269	2.519	1.381	0.214	0.530
Tb01B93I-92							

Patient	Acid Fast Sputum	ELISA Values					
		Lysate	38kD	TbRa9	TbH12	TbH4	TbRa3
Tb01B93I-9	+	2.649	0.278	0.210	0.140	0.181	1.586
Tb01B93I-14	+	>3	1.538	0.282	0.291	0.549	2.880
Tb01B93I-21	+	2.645	0.739	2.499	0.783	0.536	1.770
Tb01B93I-22	+	0.714	0.451	2.082	0.285	0.269	1.159
Tb01B93I-31	+	0.956	0.490	1.019	0.812	0.176	1.293
Tb01B93I-32	–	2.261	0.786	0.668	0.273	0.535	0.405
Tb01B93I-52	–	0.658	0.114	0.434	0.330	0.273	1.140
Tb01B93I-99	–	2.118	0.584	1.62	0.119	0.977	0.729
Tb01B94I-130	–	1.349	0.224	0.86	0.282	0.383	2.146
Tb01B94I-131	–	0.685	0.324	1.173	0.059	0.118	1.431
AT4-0070	Normal	0.072	0.043	0.092	0.071	0.040	0.039
AT4-0105	Normal	0.397	0.121	0.118	0.103	0.078	0.390
3/15/94-1	Normal	0.227	0.064	0.098	0.026	0.001	0.228
4/15/93-2	Normal	0.114	0.240	0.071	0.034	0.041	0.264
5/26/94-4	Normal	0.089	0.259	0.096	0.046	0.008	0.053
5/26/94-3	Normal	0.139	0.093	0.085	0.019	0.067	0.01

Based on cut-off values obtained from receiver-operator curves, TbRa3 detected 23 out of 27 positive sera, TbRa9 detected 22 out of 27, [illegible]

[illegible] recombinant antigens detected positive sera that were not detected by the 38 [illegible]

The reactivity of the recombinant antigen TbRa11 with sera from *M. tuberculosis* patients shown to be negative for the 38 kD antigen, as well as with sera from PPD positive and normal donors, was determined by ELISA as described above. The results are shown in Figure 6 which indicates that TbRa11, while being negative with sera from PPD positive and normal donors, detected sera that were negative with the 38 kD antigen. Of the thirteen 38 kD negative sera tested, nine were positive with TbRa11, indicating that this antigen may be reacting with a sub-group of 38 kD antigen negative sera. In contrast, in a group of 38 kD positive sera where TbRa11 was reactive, the mean OD 450 for TbRa11 was lower than that for the 38 kD antigen. The data indicate an inverse relationship between the presence of TbRa11 activity and 38 kD positivity.

The antigen TbRa2A was tested in an indirect ELISA using initially 50 µl of serum at 1:100 dilution for 30 minutes at room temperature followed by washing in PBS Tween and incubating for 30 minutes with biotinylated Protein A (Zymed, San Francisco, CA) at a 1:10,000 dilution. Following washing, 50 µl of streptavidin-horseradish peroxidase (Zymed) at 1:10,000 dilution was added and the mixture incubated for 30 minutes. After washing, the assay was developed with TMB substrate as described above. The reactivity of TbRa2A with sera from *M. tuberculosis* patients and normal donors in shown in Table 4. The mean value for reactivity of TbRa2A with sera from *M. tuberculosis* patients was 0.444 with a standard deviation of 0.309. The mean for reactivity with sera from normal donors was 0.109 with a standard deviation of 0.029. Testing of 38 kD negative sera (Figure 7) also indicated that the TbRa2A antigen was capable of detecting sera in this category.

TABLE 4

REACTIVITY OF TBRA2A WITH SERA FROM *M. TUBERCULOSIS* PATIENTS AND FROM NORMAL DONORS

[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]

Tb91	TB	0.393
Tb92	TB	0.401
Tb93	TB	0.232
Tb94	TB	0.333
Tb95	TB	0.435
Tb96	TB	0.284
Tb97	TB	0.320
Tb99	TB	0.328
Tb100	TB	0.817
Tb101	TB	0.607
Tb102	TB	0.191
Tb103	TB	0.228
Tb107	TB	0.324
Tb109	TB	1.572
Tb112	TB	0.338
DL4-0176	Normal	0.036
AT4-0043	Normal	0.126
AT4-0044	Normal	0.130
AT4-0052	Normal	0.135
AT4-0053	Normal	0.133
AT4-0062	Normal	0.128
AT4-0070	Normal	0.088
AT4-0091	Normal	0.108
AT4-0100	Normal	0.106
AT4-0105	Normal	0.108
AT4-0109	Normal	0.105

The reactivity of the recombinant antigen (g) (SEQ ID NO: 60) with sera from *M. tuberculosis* patients and normal donors was determined by ELISA as described above. Figure 8 shows the results of the titration of antigen (g) with four *M. tuberculosis* positive sera that were all reactive with the 38 kD antigen and with four donor sera. All four positive sera were reactive with antigen (g).

The reactivity of the recombinant antigen TbH-29 (SEQ ID NO: 137) with

donors and with a pool of sera from *M. tuberculosis* patients. The mean OD 450 was demonstrated to be higher with sera from *M. tuberculosis* patients than from normal donors, with the mean OD 450 being significantly higher in the indirect ELISA than in the direct ELISA. Figure 11 is a titration curve for the reactivity of recombinant TbH-33 with sera from *M. tuberculosis* patients and from normal donors showing an increase in OD 450 with increasing concentration of antigen.

The reactivity of the recombinant antigens RDIF6, RDIF8 and RDIF10 (SEQ ID NOS: 184-187, respectively) with sera from *M. tuberculosis* patients and normal donors was determined by ELISA as described above. RDIF6 detected 6 out of 32 *M. tuberculosis* sera and 0 out of 15 normal sera; RDIF8 detected 14 out of 32 *M. tuberculosis* sera and 0 out of 15 normal sera; and RDIF10 detected 4 out of 27 *M. tuberculosis* sera and 1 out of 15 normal sera. In addition, RDIF10 was found to detect 0 out of 5 sera from PPD-positive donors.

EXAMPLE 7

PREPARATION AND CHARACTERIZATION OF *M. TUBERCULOSIS* FUSION PROTEINS

A fusion protein containing TbRa3, the 38 kD antigen and Tb38-1 was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR in order to facilitate their fusion and the subsequent expression of the fusion protein TbRa3-38 kD-Tb38-1. TbRa3, 38 kD and Tb38-1 DNA was used to perform PCR using the primers PDM-64 and PDM-65 (SEQ ID NO: 141 and 142), PDM-57 and PDM-58 (SEQ ID NO: 143 and 144), and PDM-69 and PDM-60 (SEQ ID NO: 145-146), respectively. In each case, the DNA amplification was performed using 10 µl 10X Pfu buffer, 2 µl 10 mM dNTPs, 2 µl each of the PCR primers [illegible]

68°C for 15 sec and 72°C for 3 min, and finally by 72°C for 4 min. For Tb38-1 denaturation at 94°C for 2 min was followed by 10 cycles of 96°C for 15 sec, 68°C for 15 sec and 72°C for 1.5 min, 30 cycles of 96°C for 15 sec, 64°C for 15 sec and 72°C for 1.5, and finally by 72°C for 4 min.

The TbRa3 PCR fragment was digested with NdeI and EcoRI and cloned directly into pT7^L2 II. 1 vector using NdeI and EcoRI sites. The 38 kD PCR fragment was digested with Sse8387I, treated with T4 DNA polymerase to make blunt ends and then digested with EcoRI for direct cloning into the pT7^L2Ra3-1 vector which was digested with StuI and EcoRI. The 38-1 PCR fragment was digested with Eco47III and EcoRI and directly subcloned into pT7^L2Ra3/38kD-17 digested with the same enzymes. The whole fusion was then transferred to pET28b using NdeI and EcoRI sites. The fusion construct was confirmed by DNA sequencing.

The expression construct was transformed to BLR pLys S *E. coli* (Novagen, Madison, WI) and grown overnight in LB broth with kanamycin (30 μg/ml) and chloramphenicol (34 μg/ml). This culture (12 ml) was used to inoculate 500 ml 2XYT with the same antibiotics and the culture was induced with IPTG at an OD560 of 0.44 to a final concentration of 1.2 mM. Four hours post-induction, the bacteria were harvested and sonicated in 20 mM Tris (8.0), 100 mM NaCl, 0.1% DOC, 20 μg/ml Leupeptin, 20 mM PMSF followed by centrifugation at 26,000 X g. The resulting pellet was resuspended in 8 M urea, 20 mM Tris (8.0), 100 mM NaCl and bound to Pro-bond nickel resin (Invitrogen, Carlsbad, CA). The column was washed several times with the above buffer then eluted with an imidazole gradient (50 mM, 100 mM, 500 mM imidazole was added to 8 M urea, 20 mM Tris (8.0), 100 mM NaCl). The eluates containing the protein of interest were then dialzyed against 10 mM Tris (8.0)

The DNA and amino acid sequences for the resulting fusion protein

procedure to that described above. The DNA sequence for the TbH9-Tb38-1 fusion protein is provided in SEQ ID NO: 151.

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and DPEP was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR and cloned into vectors essentially as described above, with the primers PDM-69 (SEQ ID NO:145 and PDM-83 (SEQ ID NO: 200) being used for amplification of the Tb38-1A fragment. Tb38-1A differs from Tb38-1 by a DraI site at the 3' end of the coding region that keeps the final amino acid intact while creating a blunt restriction site that is in frame. The TbRa3/38kD/Tb38-1A fusion was then transferred to pET28b using NdeI and EcoR1 sites.

DPEP DNA was used to perform PCR using the primers PDM-84 and PDM-85 (SEQ ID NO: 201 and 202, respectively) and 1 µl DNA at 50 ng/µl. Denaturation at 94 °C was performed for 2 min, followed by 10 cycles of 96 °C for 15 sec, 68 °C for 15 sec and 72 °C for 1.5 min; 30 cycles of 96 °C for 15 sec, 64 °C for 15 sec and 72 °C for 1.5 min; and finally by 72 °C for 4 min. The DPEP PCR fragment was digested with EcoRI and Eco72I and clones directly into the pET28Ra3/38kD/38-1A construct which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-2) are provided in SEQ ID NO: 203 and 204, respectively.

EXAMPLE 8

USE OF *M. TUBERCULOSIS* FUSION PROTEINS FOR SERODIAGNOSIS OF TUBERCULOSIS

The effectiveness of the fusion protein TbRa3-38 kD-Tb38-1, prepared as described above, in the serodiagnosis of tuberculosis infection was examined by ELISA.

The ELISA protocol was as described above in Example 6, with the fusion

the three antigens individually or in combination. Such a panel enabled the dissection of the serological reactivity of the fusion protein to determine if all three epitopes functioned with the fusion protein. As shown in Table 5, all four sera that reacted with TbRa3 only were detectable with the fusion protein. Three sera that reacted only with Tb38-1 were also detectable, as were two sear that reacted with 38 kD alone. The remaining 15 sera were all positive with the fusion protein based on a cut-off in the assay of mean negatives +3 standard deviations. This data demonstrates the functional activity of all three epitopes in the fusion protein.

TABLE 5

REACTIVITY OF TRI-PEPTIDE FUSION PROTEIN WITH SERA FROM *M. TUBERCULOSIS* PATIENTS

Serum ID	Status	ELISA and/or Western Blot Reactivity with Individual proteins			Fusion recombinant OD 450	Fusion Recombinant Status
		38kd	Tb38-1	TbRa3		
01B93I-40	TB	-	-	+	0.413	+
01B93I-41	TB	-	+	+	0.392	+
01B93I-29	TB	+	-	+	2.217	+
01B93I-109	TB	+	+	+	0.522	+
01B93I-132	TB	+	+	+	0.937	+
5004	TB	+	+	+	1.098	+
15004	TB	+	+	+	2.077	+
39004	TB	+	+	+	1.675	+
68004	TB	+	+	+	2.388	+
99004	TB	-	+	+	0.607	+
107004	TB	-	+	+	0.667	+
92004	TB	+	+	+	1.070	+
97004	TB	+	-	+	1.152	+
118004	TB	+	-	+	2.694	+
173004	TB	+	+	+	3.258	+
175004	TB	+	-	+	2.514	+
274004	TB					

289004	TB	-	-	+	0.848	+
308004	TB	-	+	-	3.338	+
314004	TB	-	+	-	1.362	+
317004	TB	+	-	-	0.763	+
312004	TB	-	-	+	1.079	+
D176	PPD	-	-	-	0.145	-
D162	PPD	-	-	-	0.073	-
D161	PPD	-	-	-	0.097	-
D27	PPD	-	-	-	0.082	-
A6-124	NORMAL	-		-	0.053	-
A6-125	NORMAL	-	-	-	0.087	-
A6-126	NORMAL	-	-	-	0.346	+
A6-127	NORMAL	-	-	-	0.064	-
A6-128	NORMAL	-	-	-	0.034	-
A6-129	NORMAL	-	-	-	0.037	-
A6-130	NORMAL	-	-	-	0.057	-
A6-131	NORMAL	-	-	-	0.054	-
A6-132	NORMAL	-	-		0.022	-
A6-133	NORMAL	-	-		0.147	-
A6-134	NORMAL	-	-	-	0.101	-
A6-135	NORMAL	-	-		0.066	-
A6-136	NORMAL	-	-		0.054	-
A6-137	NORMAL	-	-	-	0.065	-
A6-138	NORMAL	-	-	-	0.041	-
A6-139	NORMAL	-			0.103	
A6-140	NORMAL	-	-	-	0.212	-
A6-141	NORMAL		-	-	0.056	-
A6-142	NORMAL	-	-	-	0.051	-

The reactivity of the fusion protein TbI-2 with sera from *M. tuberculosis*-infected patients was examined by ELISA using the protocol described above. The results of these studies (Table 6) demonstrate that

TABLE 6

REACTIVITY OF TbF-2 FUSION PROTEIN WITH TB AND NORMAL SERA

Serum ID	Status	TbF OD450	Status	TbF-2 OD450	Status	ELISA Reactivity			
						38 kD	TbRa3	Tb38-1	DPEP
B931-40	TB	0.57	+	0.321	+	-	+	-	+
B931-41	TB	0.601	+	0.396	+	+	+	+	-
B931-109	TB	0.494	+	0.404	+	+	+	±	-
B931-132	TB	1.502	+	1.292	+	+	+	+	±
5004	TB	1.806	+	1.666	+	±	+	+	-
15004	TB	2.862	+	2.468	+	+	+	+	-
39004	TB	2.443	+	1.722	+	+	+	+	-
68004	TB	2.871	+	2.575	+	+	+	+	-
99004	TB	0.691	+	0.971	+	-	+	+	-
107004	TB	0.875	+	0.732	+	-	±	+	-
92004	TB	1.632	+	1.394	+	+	±	±	-
97004	TB	1.491	+	1.979	+	+	±	-	+
118004	TB	3.182	+	3.045	+	+	±	-	-
173004	TB	3.644	+	3.578	+	+	+	+	-
175004	TB	3.332	+	2.916	+	+	+	-	-
274004	TB	3.696	+	3.716	+	-	+	-	+
276004	TB	3.243	+	2.56	+	-	-	+	-
282004	TB	1.249	+	1.234	+	+	-	-	-
289004	TB	1.373	+	1.17	+	-	+	-	-
308004	TB	3.708	+	3.355	+	-	-	+	-
314004	TB	1.663	+	1.399	+	-	-	+	-
317004	TB	1.163	+	0.92	+	+	-	-	-
312004	TB	1.709	+	1.453	+	-	+	-	-
380004	TB	0.238	-	0.461	+	-	+	-	+
451004	TB	0.18	-	0.2	-	-	-	-	+
478004	TB	0.188	-	0.469	+	-	-	-	+
410004	TB	0.384	+	2.392	+	+	-	-	+
411004	TB	0.306	+	0.874	+	-	+	-	+
421004	TB	0.357	+	1.456	+	-	+	-	+
528004	TB	0.047	-	0.196	+	-	-	-	+
A6-87	Normal	0.094		0.063		-	-	-	-
A6-88	Normal	0.214		0.19		-	-	-	-
A6-89	Normal	0.248		0.125		-	-	-	-
A6-90	Normal	0.179	-	0.206	+	-	-	-	-
A6-91	Normal	0.135	-	0.151		-	-	-	-
A6-92	Normal	0.064		0.097		-	-	-	-
A6-93	Normal	0.072	-	0.098		-	-	-	-
A6-94	Normal	0.072		0.064		-	-	-	-
A6-95	Normal	0.125		0.150		-	-	-	-

One of skill in the art will appreciate that the order of the individual antigens within the fusion protein may be changed and that comparable activity would be expected provided each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the construction of fusion proteins.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANTS: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonia
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.

(ii) TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

(iii) NUMBER OF SEQUENCES: 209

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SEED and BERRY LLP
(B) STREET: 6300 Columbia Center, 701 Fifth Avenue
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: USA
(F) ZIP: 98104-7092

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: Patent[n Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 01-OCT-1997
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/027,757
(B) FILING DATE: 04-OCT-1996

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Reed, Dianne E.
(B) REGISTRATION NUMBER: 32,481
(C) REFERENCE/DOCKET NUMBER: 210.1.1

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (206) 622-4900
(B) TELEFAX: (206) 682-6031

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGAGGCACCG GTAGTTTGAA CCAAACGCAC AATCGACGGG CAAACGAACG GAAGAACACA 60

ACCATGAAGA TGGTGAAATC CATCGCCGCA GGTCTGACCG CCGCGGCTGC AATCGGCGCC 120

GCTGCGGCCG GTGTGACTTC CATCATGGCT GCCGGCCCCG TCGTATACCA GATGCAGCCG 180

GTCGTCTTCG GCGCGCCACT GCCGTTGGAC CCGGCATCCG CCTCTGACGT CCCGACCGCC 240

GCCCAGTTGA CCAGCTGCCT CAACAGCCTC GCCGATCCCA ACGTGTCGTT TGCGAACAAG 300

GGCAGTCTGG TCGAGGGCGG CATCGGGGGC ACCGAGGCGC GCATCGCCGA CCACAAGCTG 360

AAGAAGTCCG CCGAGCACGG GGATCTGCCG CTGTCGTTCA GCGTGACGAA CATCCAGCCG 420

GCGGCCGCCG GTTCGGCCAC CGTCGACGTT TCCGTCTCGG GTCCGAAGCT CTCGTCGCCG 480

GTCACGCAGA ACGTCACGTT CGTGAATCAA GGCGGCTGGA TGCTGTCACG CGCATCGGGG 540

ATGGAGTTGC TGCAGGCCGC AGGGNAACTG ATTGNCGGGC CGGNTTCAGC CCGCTGTTCA 600

GCTACGCCGC CCGCCTGGTG ACGCGTCCAT GTCGAACACT CGCCGCGTGA GCACCGTGCG 660

GTNTGCCCAG GCNCCCACGC ACCGCCCGGT GTAAGCCGTC GTCGAGATAG GTGGTGNCTC 720

GNTACCACNG ANTACCTCCC NNTCNNCNTT TCTCANTCGT GNATGA 766

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 750 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

TTTCTCGACG ACGTGACCGT GAGCCGTCGC CATGCTGAAT TCCCGTTGGA AAACAACGAA 300

TTCAATGTCG TCGATGTCGG GAGTCTCAAC GGCACCTACC TCAACCGCGA GCCCGTGGAT 360

TCGGCGGTGC TGGCGAACGG CGACGAGGTC CAGATCGGCA AGCTCCGGTT GGTGTTCTTG 420

ACCGGACCCA AGCAAGGCGA GGATGACGGG AGTACCGGGG CCCCGTGAGC GCACCCGATA 480

GCCCCGCGCT GGCCGGGATG TCGATCGGGG CGGTCCTCCG ACCTGCTACG ACCGGATTTT 540

CCCTGATGTC CACCATCTCC AAGATTCGAT TCTTGGGAGG CTTGAGGGTC NGGGTGACCC 600

CCCCGGGGGC CTCATTCNGG GGTNTCGGCN GGTTTCACCC CNTACCNACT GCCNCCTGGN 660

TTGCNAATTC NTTCTTCNCT GCCTNNAAAA GGACCNTTAN CTTGCCGCIN GAAANGGTNA 720

TCCNGGGCCC NTCCTNGAAN CCCCNTCCCC CT 752

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 813 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GATATGCATT ACCATCACTA TCATACTTCT AATGGTTAAT CCTGTCGGGG GCGTCTAGCA 60

CCAGCCGACA CCGGGCTGGA CGGATTCGCT AGTTTGACTC GCTCAGGTA TGGTCGTCAG 120

CAGTCCGATG CCCTATGTTT GTCGTCGACT TAGATATCGG GCAATTCAA TCTCCCCACT 180

[illegible] 240

[illegible] 300

[illegible] 360

[illegible] 420

[illegible] 480

[illegible] 540

GAGCAACGCA GACCGGGACA ACWGGTATCG ATAGCGGCCN AATGCCGGGT TGGAACCCNG 780

TGAAATTATC ACAACTTCGC AGTCACNAAA NAA 813

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 447 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGGTATGAAT ACGGCCGGGT CCGATAACTT CGAGTTGTCT CAGGGTGGGC AGGGATTCGC 60

CATTCCGATC GGGCAGGCGA TGGCGATCGC GGGCCAGATC CGATCGCGTG GCGCGTCAGC 120

CACCGTTCAT ATCGGGCCTA CCGGCTTCCT CGGCTTGGGT GTTGTCGACA ACAACGGCAA 180

CGGCGCACGA GTCCAACCCG TGGTTGGGAG CGATCGGGCT GCAAGTCTCG GCATCTCCAC 240

CGGCGACGTG ATCACCGCGG TCGACGGCGC TCCGATCAAC TCGGCCACCG CGATGGCGGA 300

CGCGCTTAAC CGGCATCATT CGCGTGAACT CATTCGGTTG AACTGGCAAA CGAAGTCGGG 360

CGGCACGCTC ATACGGAACT TGATTGTCTC CGAGGACCCG CGCAGTTTAT TTCGTGATGG 420

ATATTATTTT CGGTCTCTTT AATTCGA 447

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible]
(B) TYPE: [illegible]
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

CCGGCGACGG NGAGCGCCGG AATGGCGCGA GTGACGACGT GGNCAGTCAT GCCCAGNGTG 240

ATCCAATCAA CCTGNATTCG GNCTGNGGGN CCATTTGACA ATCGAGGTAG TGAGCGCAAA 300

TGAATGATGG AAAACGGGNG GNGACGTCCG NTGTTCTGGT GGTGNTAGGT GNCTGNCTGG 360

NGTNGNGGNT ATCAGGATGT TCTTCGNCGA AANCTGATGN CGAGGAACAG GGTGTNCCCG 420

NNANNCCNAN GGNGTCCNAN CCCNNNNTCC TCGNCGANAT CANANAGNCG NTTGATGNGA 480

NAAAAGGGTG GANCAGNNNN AANTNGNGGN CCNAANAANC NNNANNCNNG NNAGNTNGNT 540

NNNTNTTNNC ANNNNNNNTG NNGNNGNNCN NNNCAANCNN NTNNNNGNAA NNGGNTTNTT 600

NAAT 604

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 633 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTGCANCTCG AACCACTCTA CTAAACGAA CAAAAGCTGC AGTTCACTC TGGTGGCGGC 60

CCTTTAGAA CTACTNATN YTTCTCCTT TANAATTC CYACGAGTAT TANGACAGTC 120

TAANCTCCT CTTANCTTN TNAATAAT TANAATTC TCTTATTCA TATTACGAA 180

CCCCCCGAA CCCTCACCT CAACTCCC CATCTTCTA ANCCCTCT CCCCCCATA 240

[illegible] 300

[illegible] 360

[illegible] 420

[illegible] 480

[illegible] 540

[illegible] 600

[illegible] 633

(A) LENGTH: 1362 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGACGACGAC GGTGCCGCAG AGCGGGTGCG AACGGCGATC GACGCGGCCC TGGCCAGAGT 60

CGGGACCACC CAGGAGGGAA TCGAATCATG AAATTTGTCA ACGATATTGA GCCCGTCGCG 120

CCCCGCCGAG CCGGCGGCGC GTTCGCTGAG GTCTATGCCG ACGCCCCGCT CGAGTTCGGC 180

CGGCTGCTCG AGCCCGTTGC CATGCTGTCC CCGGACGAGG GACTGCTCAC CGCCGGCTGG 240

GCGACGTTGC GCGAGACACT GTTGGTCGGC CAGGTGCCGC GTGGCCGCAA GGAAGCGGTC 300

GCTGCCGCCG TCGCGCGCCAG CCTGCGCTGC CCCTGGTGCG TCGACGCACA CACCACCATG 360

CTGTACGCGG CAGGCCAAAC CGACACCGCC GCGCCGATCT TGGCCGGCAC AGCACCTGCC 420

GCCGGTGACC CGAACGCGCC GTATGTGGCG TGGGCGGCAG GAACTGGGAC ACCGGCGGGA 480

GCACGGGCAC CGTTCGGCCC AGATGTCGCC GCCGAATACC TGGGCACCGG GGTGCAATTC 540

GACTTCATCG CACGCCTGCT CTCGTCGCTC CTGCACGAAA TCTTCCTCGC CGGGGGCCCG 600

CGCCCGAAC AGCTCATCGA CGACGCCGCA GAACTGCTCT TCTCCCGCAA GTTCGCCGCG 660

GAGCATCCGG CGGGCCGGTC CATCCGCGAC CTGAAGCGCG AACTGCTCAT CGATCTG 720

GCATGGCAA CACCGTCCGA CCTATACGA ATTCGTTCC CTTCGCTAT CTGCCACTG 780

GACACCCGG CGCACCTGCT CCTACGACT TTCACGTCC TCGACGCCC CTTCGCTCC 840

GGACCGGACG GTCACCGGGG GTCACCCTGC GCGGCCAAGG AA 1362

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1458 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCGACGACCC CGATATGCCT CGTACCGTAA CGAAAGCCGT CGCCGACGTA GTCGGGCGCG 60

GTATCGTTCC CGTTGAGGAA ATTCAGGACT GCGTGGAGGC CTGGCTGGGG GAAGCCGGTC 120

TGGATGACGT GGCCCGTGTT TACATCATCT ACCGGCAGCG GCGCGCCGAG CTGCGGACGG 180

CTAAGGCCTT GCTCGGCGTG CGGGACGAGT TAAAGCTGAG CTTGGCGGCC GTGACGGTAC 240

TGCGCGAGCG CTATCTGCTG CACGACGAGC AGGGCCGGCC GGCCGAGTCG ACCGGCGAGC 300

TGATGGACCG ATCGGCGCGC TGTGTTGCGG CGGCCGAGGA CCAGTATGAG CCGGGCTCGT 360

CGAGGCGGTG GGCCGAGCGG TTCGCCAGGC TATTACGCAA CCTCGAATTT GTGCCGAATT 420

CGCCCACGTT GATGAACTTT GCTACTGACT TTGACTCTCT CGTCGGCTCT TTTGTTCTGC 480

TGATTGAGGA TTCGTTCTAA TTGACTTTCT CGACTTCTGG ACAAGCCCCT GAGCTGCAGC 540

GGGCTGGAGG CGGTACTGTA TATTCGTTCA CTACTTCTCG ACTCCGCCGG TATCGGGTGG 600

CCTCCAAGGG CGGCACGGTC AGCTGACGCT CTTTTCT ACCGCTTCC GACATCGCCC 660

CGGCCCGCGC CACCCGCAAG ATCGGGCTAG GAGTCATGGG TTTGGCGGAA CTGCTTGCCG 1260

CACTGGGTAT TCCGTACGAC AGTGAAGAAG CCGTGCGGTT AGCCACCCGG CTCATGCGTC 1320

GCATACAGCA GGCGGCGCAC ACGGCATCGC GGAGGCTGGC CGAAGAGCGG GGCGCATTCC 1380

CGGCGTTCAC CGATAGCCGG TTCGCGCGGT CGGGCCCGAG GCGCAACGCA CAGGTCACCT 1440

CCGTCGCTCC GACGGGCA 1458

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 862 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ACGGTGTAAT CGTGCTGGAT CTGGAACCGT TTGGCCGCGT ACGTACCGAC ATGTACTGGC 60

GGCGCAGGGG GCTGGCCTTG GGCATCGTGG TCGTCGTAGT CGGGATCGGG CTGGCCATCG 120

TCATCCCGTT CTTGACAAGC AGCGCCAGTG CGAAACCGGT CAACGCCGAC AAGCCGGCCT 180

CGGCTCAGAG CATTCGTGTC TACCCGGTAG CCAACGTAGT CCAGCCCGCC GGGCAAACCG 240

AACTTAACTC CGTCCGTCTC GTCCTCAAGG GTAAAACCT TAACGACGCC AGCTTAACG 300

TTCCCGTTCA CTCCCCGTCG TTTTTAAGGT AATTCGACTA TTTTTTGAT TGACTTCG 360

CCGTTAAAGG TTTCATAAAC CCCTTTAGC ATTCTTCTT CGATAGCTAA CGTTTCATTA 420

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 622 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTGATCAGCA CCGGCAAGGC GTCACATGCC TCCCTGGGTG TGCAGGTGAC CAATGACAAA 60

GACACCTTTG GTGCGAACAT CGTCGAAGTA ATGGCCGGTG TGCTGCCGCT GAACGCTGGA 120

GTTCCGAAGG GTTCGTTGGT AACTAAGTTT GACGACCCGC CGATCAAAAG TGCGGACGCG 180

TTGGTTGCGC CGTGCGGTTC CAAAGCGCCT GGCGCCACGG TGCCGCTAAC CTTTCAGGAT 240

CCGTCGGGGA GTAGCGGCAC AGTGCAAGTC ACCCTCGGCA AGGCGGAGCA GTGATGAAGG 300

TCGCCGCTGA GTGTTCAAAG CTCGGATATA CGGTGGCACC CATGGAACAG CGTGCGGAGT 360

TGGTGGTTTG GCGGGACTT GTCGTCGTCG TTGACGATCG CACGGCGCAC CGCGATGAAG 420

ACCACAGCTT TCGTTTTTTC ACCGACGTTT CAGCGAACTG CGGTTTCTTT GTCGACGGCG 480

TGGTGGCGGT GTCGTTCGA GACTTGGAAA TCAAAATGGT CATGAACATA CGGGTGATCG 540

GCGGGGTGGA CCTGGTGGTG TTCGTCGGCA CAGTCGGCAT GACGATTATC GATGTCACCC 600

TGAAACCAC CTGGACATT CT 622

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

GCCTACGTGC GATCGTGCCC GGGCTACACG TTGGACTACA ACGCCAACGG GTCCGGTGCC 240

GGGGTGACCC AGTTTCTCAA CAACGAAACC GATTTCGCCG GCTCGGATGT CCCGTTGAAT 300

CCGTCGACCG GTCAACCTGA CCGGTCGGCG GAGCGGTGCG GTTCCCCGGC ATGGGACCTG 360

CCGACGGTGT TCGGCCCGAT CGCGATCACC TACAATATCA AGGGCGTGAG CACGCTGAAT 420

CTTGACGGAC CGACTACCGC CAAGATTTTC AACGGCACCA TCACCGTGTG GAATGATCCA 480

CAGATCCAAG CGCTGAACTC CGGCACCGAC CTGCCGCCAA CACCGATTAG CGTTATCTTC 540

CGCAGCGACA AGTCCGGTAC GTCGGACAAC TTCCAGAAAT ACCTCGACGG TGTATCCAAC 600

GGGGCGTGGG CGAAAGGCGC CAGCGAAACG TTCAGCGGGG GCGTCGGCGT CGGCGCCAGC 660

GGGAACAACG GAACGTCCGC CCTACTGCAG ACGACCGACG GTTCGATCAC CTACAAGGAG 720

TCGTCGTTTG CGGTGGGTAA GCAGTTGAAC ATGGCCCAGA TCATCACGTC GGCGGGTCCG 780

GATCCAGTGG CGATCACCAC CGAGTCGGTC CGTAAGACAA TCGCCGGCGC CAAGATCATT 840

CGACAAGGCA ACGACCTGGT ATTGGACACG TCGTCGTTCT ACAGACCCAC CCAGCCTGGT 900

TCTTACCCGA TCGTGCTGGC GACCTATGAG ATCGTCTGCT CGAAATACCC GGATGCGACG 960

ACCGGTACTG CGGTAAGGGC GTTTATCAAA GCCGCCATTG GTCCAGGCCA AGAAGGCCTG 1020

GACCAATACC CTTCCATTCC GTTCCGTAAA TCGTTTCAAG TAAAATTCTA CGCCGCGTC 1080

AAACGTACTT CTTGACCTAA CGAAGGAAT TCGACGCTGA CGACGCCGT TCTCGAGCTA 1140

GCTTCCCAAT TTCCCCTCGA TACGTATTC CCCCTATAC CCCACCCCGA 1188

(2) INFORMATION FOR SEQ ID NO:2:

GGTTGCTCCA AGCGGTGGCC GCCGACGGCC GCATCCACAC CACGTTCAAC CAGACGATCG 240

CCGCGACCGG CCGGCTCTCC TCGACCGAAC CCAACCTGCA GAACATCCCG ATCCGCACCG 300

ACGCGGGCCG GCGGATCCGG GACGCGTTCG TGGTCGGGGA CGGTTACGCC GAGTTGATGA 360

CGGCCGACTA CAGCCAGATC GAGATGCCGA TCATGGGGCA CCTGTCCGGG GACGAGGGCC 420

TCATCGAGGC GTTCAACACC GGCGAGGACC TGTATTCGTT CGTCGCGTCC CGGGTGTTCG 480

GTGTGCCTAT CGACGAGGTC ACCGGCGAGT TGCGGCGCCG GGTCAAGGCG ATGTCGTACG 540

GGCTGGTTTA CGGGTTGAGC GACTACGGCC TGTCGCAGCA GTTGAAAATC TCCACCGAGG 600

AAGCGAACGA GCAGATGGAC GCGTATTTCG CCCGATTCGG CGGCGTCGCC GACTACCTGG 660

ACGCGGTAGT CGAGCGGGCC CGCAAGGACG GCTACACCTC GACGGTGCTG GGCCGTCGCC 720

GCTACCTGCC CGACCTGGAC AGCAGCAACC GTCAAGTGCG GGAGGCCGCC GAGTGGGCGA 780

CGCTGAACGC GCCCATCCAG GGCAGCGCGG CCGACATCAT CAAGGTGGCC ATGATCCAGG 840

TCGACAAGGC GCTCAACGAG GCACAGCTGG CGTCGCGCAT GCTGCTGCAG GTCCACGACG 900

AGCTGCTGTT CGAAATCGCC GCCCCGTCAA TCGAGCAGGT CGAGGCGCTG GTGCGCGACA 960

AGATGGAGGG CGCTTACCCG CTGGACCTGC CGCTGCAGGT CTCGGTGGGC TACGCCGCGA 1020

GCTGGGACGC GCACGGCCAC TGAGTGCCGA CGGTGCATTC GCGCAGCAAA TTCGGCGATT 1080

TTTCGGGCGT AGTTAGGCAC CGGCTGAAC CGGATCGAT TTTTTCAGCT GTGTACCGCT 1140

CGATTAGCCT CGTCA 1155

(2) INFORMATION FOR SEQ ID NO:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

GGCCCTCGGA TGCCGATTCC GCAGCTTTCT GGTGGCGACGG CTGGCCCTTC GAGCACGGAC 240

ATCGAGAACT CTCGGGGTTC GGCCAACGTT ATCTCAGTGG AATCTCAGTC AACGCGCGCA 300

ACCTAGTTGT GCAGTTACTG TTGAAAGCCA CACCCATGCC AGTCCACGCA TGGCCAAGTT 360

GGCCCGAGTA GTGGGCCTAG TACAGGAAGA GCAACCTAGC GACATGACGA ATCACCCACG 420

GTATTCGCCA CCGCCGCAGC AGCCGGGAAC CCCAGGTTAT GCTCAGGGGC AGCAGCAAAC 480

GTACAGCCAG CAGTTCGACT GGCGTTACCC ACCGTCCCCG GCCCCGCAGC CAACCCAGTA 540

CCTTCAACGC TATGAGGCGT TGGGTGGTAC CCGGCCGGGT CTGATATTTG GCGTGATTCC 600

GACCATGACG CCTCGTCCTG GGATGGTTCG CCAACGACCT CTTGCAGGCA TGTTGGCCAT 660

CGGCGACCTG ACGATAGCGG TGGTGTCGGC GGGCATCGGC GGCGGCGACG CATCGCTGGT 720

CGGGTTCAAC CGGCCACCGG CGGTAAACAG CGGCGGCCCA GTGGCTGTTA CGGTGGCGCC 780

AAGCATCCGC GCAGCAAACA TGCCGCCGGG GTCGGTCGAA CAGGTGGCGG CCAAGGTGGT 840

GCGCAGTGTC GTCATGTTGG AAACCGATCT GGGCCGCCAG TCGGAGGAAA GCTCGGGCAT 900

CATTCGTGCT GCCCAGGGGC TGATCTTGAC CAACAACCAC GTCATCGCCG CGGCCGCCAA 960

GCGTCCGCCG GGCAGTCCGC CGCCGAAAAC GACGGTAACC TTCTCTGACG GGCCGACCGC 1020

ACCCTTCATC GTGGTGGGGC CTGACCCCAC CAGTGATATC GCCGTCGTCG GTGTTCAGGG 1080

CATCTTCATT CTCACCCGGA TGTCCCTGGG TTCGTCCTCG GACCTGAGCT TCGGTCAGCC 1140

GCCTTTGATT ATCGGTCGT CGCTCGTTT GGAGGGCAAT CTCAACCTCT GGATCGTCAG 1200

TCTCTTAAT GTCAGTGT CGACGACT CAACCCCCT AACTGGAA TTCCCTGA 1260

CGCCATTCAG ACCCAGCCG CGATCAACGC GTGTAACTCC GCCAGCGTTT TGTTTGAACAT 1320

GA TGAAA GA 1380

GAGCGT GA 1440

GGCGCGCCGC 1500

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1058 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCCACCGCG CTGGCGGCCG CTCTAGAACT AGTGGATCCC CCGGGCTGCA GGAATTCGGC 60

ACGAGGATCC GACGTCGCAG GTTGTCGAAC CCGCCGCCGC GAAGTATCCG GTCCATCCCT 120

AGCTCGGCGA CGGCGAGCGC CGGAATGGCG CGAGTGAGGA GGCGGGCAAT TTGGCGGGGC 180

CGGGCGACGG CGAGGGCGGC AATTGGGCGA GTGAGGAGGC GGGCAGTCAT GCCCAGGGTG 240

ATCCAATCAA CCTGCATTCC GCCTGCGGGC CCATTTGACA ATCGAGGTAC TGAGCGCAAA 300

TGAATGATGG AAAACGGGCG GTGACGTCCG CTGTTCTGGT GGTGCTAGGT GCCTGCCTGG 360

CGTTGTGGCT ATCAGGATGT TCTTCGCCGA AACCTGATGC CGAGGAACAG GGTGTTCCCG 420

TGAGCCCGAC GGCGTCCGAC CGCCGCGCTC TTGCCGAGAT CAGGCAGTCG CTTGATGCGA 480

CAAAAGCGTT GACCAGCGTC CACGTAGCGC TTGAACAACT CGGAAAGTTC GACAGCTTGC 540

TGCGTATTAC CAGTGCCGAC GTTGACGTTC CGGACAATCC GATCGCGGCA AAGCGCGTAT 600

CGACTTATAA CGACGAGTAC GATTTGCTTC TTGCGTTAAA AGAGACAAAT AGTTCTGCGA 660

AACTGTTTGA CGACTGGATT AATCTTGCGC TTGATTCTTA ATGTAACT TCAGCGGTGC 720

TCGATCCTGT TCTTCAGGTC AGCAATTGCT TGTTTGATTC TAGGAATTCT TAACGCAAAG 780

[illegible] 840

[illegible] 900

[illegible] 960

[illegible] 1020

[illegible] 1058

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAATTCGGCA CGAGAGGTGA TCGACATCAT CGGGACCAGC CCCACATCCT GGGAACAGGC 60

GGCGGCGGAG GCGGTCCAGC GGGCGCGGGA TAGCGTCGAT GACATCCGCG TCGCTCGGGT 120

CATTGAGCAG GACATGGCCG TGGATAGCGT CGGCAAGATC ACCTACCGCA TCAAGCTCGA 180

AGTGTCGTTC AAGATGAGGC CGGCGCAACG GCGCTAGCAC GGGCCGGCGA GCAAGACGCA 240

AAATCGCACG GTTTGCGGTT GATTCGTGCG ATTTTGTGTC TGCTCGCCGA CGCCTACCAG 300

GTGGCGCTCA GCTACGCGTG CTGTAGTATC CAGGCGTGCA TCGCGATTCC GGCGGCCACG 360

CCGGAGTTAA TGCTTCGCGT CGACCCGAAC TGGGCGATCC GCCGCNGAGC TGATCGATGA 420

CCGTGGCCAG CCCGTCGATG CCCGAGTTGT CCGAGGAAAC GTCCTGCCAG GCCGGTAGGA 480

AGNGTCGGTA GGCGGCGGTG CTGACCGGCT CTGCCTGCGC GCTCAGTGCG GCCAGCGAGC 540

GG 542

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 814 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GTTTGCCGCC AATATTCGGC GGGCACCGCC AGATCCGCCG GGGCCACCAT TGCCGCCGGG 480

CACCGAAACA ACAGCCCAAC GGTGCCGCCG GCCCCGCCGT TTGCCGCCAT CACCGGCCAT 540

TCACCGCCAG CACCGCCGTT AATGTTTATG AACCCGGTAC CGCCAGCGCG GCCCCTATTG 600

CCGGGCGCCG GAGNGCGTGC CCGCCGGCGC CGCCAACGCC CAAAAGCCCG GGGTTGCCAC 660

CGGCCCCGCC GGACCCACCG GTCCCGCCGA TCCCCCCGTT GCCGCCGGTG CCGCCGCCAT 720

TGGTGCTGCT GAAGCCTTA GCGCCGGTTC CGCGGGTTCC GCCGGTGGCG CCNTGGCCGC 780

CGGCCCCGCC GTTGCCGTAT ACCCACCCGC CGTTGGCGCC GTTGCCGCCA TTGCCGCCAT 840

TGCCGCCGTT GCCGCCATTT CGGCCGTTCC CGTTGCCACC GCCGGNTTGG CCGCCGGCGC 900

CGCCGGCGGC CGC 913

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1872 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GACATTGAAC GGCTTGATCC AGTTCGATGC CGCAATCCAG CCCGGTGATT CGGGCGGGCC 720

CGTCGTCAAC GGCCTAGGAC AGGTGGTCGG TATGAACACG GCCGCGTCCG ATAACTTCCA 780

GCTGTCCCAG GGTGGGCAGG GATTCGCCAT TCCGATCGGG CAGGCGATGG CGATCGCGGG 840

CCAAATCCGA TCGGGTGGGG GGTCACCCAC CGTTCATATC GGGCCTACCG CCTTCCTCGG 900

CTTGGGTGTT GTCGACAACA ACGGCAACGG CGCACGAGTC CAACGCGTCG TCGGAAGCGC 960

TCCGGCGGCA AGTCTCGGCA TCTCCACCGG CGACGTGATC ACCGCGGTCG ACGGCGCTCC 1020

GATCAACTCG GCCACCGCGA TGGCGGACGC GCTTAACGGG CATCATCCCG GTGACGTCAT 1080

CTCGGTGAAC TGGCAAACCA AGTCGGGCGG CACCGGTACA CCGAACGTGA CATTGGCCGA 1140

GGGACGGCCG GCCTGATTTG TCGCCGATAC CACCTGCCGG CCGGCCAATT GGATTGGCGC 1200

CAGCCGTGAT TGCCGCGTGA GCCGCCGAGT TCCGTCTCCC GTGCGCGTGG CATTGTGGAA 1260

GCAATGAACG AGGCAGAACA CAGCGTTGAG CACCCTCCCG TGCAGGGCAG TTACGTCGAA 1320

GGCGGTGTGG TCGAGCATCC GGATGCCAAG GACTTCGGCA GCGCCGCCGC CCTGCCGGCC 1380

GATCCGACCT GGTTTAAGCA CGCCGTCTTT TACGAGGTGC TGGTCCGGGC GTTCTTCGAC 1440

GCCAGCGCGG ACGGTTCCGN CGATCTGCGT GGACTCATCG ATCCCCTCGA CTACCTGCAG 1500

TGGCTTGGCA TCGACTGCAT CTGTTGCCGC CGTTCCTACG ACTCACCGCT GCCGGACGGC 1560

GGTTACGACA TTCGCGACTT CTACAAGGTG CTGCCCGAAT TCGGCACCGT CGACGATTTC 1620

GTCGCCCTGG TCGACA[illegible] TCACC[illegible] [illegible] [illegible] [illegible]TGCTGATC 1680

AATCACACCT CGGAGTC[illegible] [illegible]TGCTTT CA[illegible] [illegible] [illegible] [illegible]

TACGGTGACT ATTACG[illegible] [illegible] [illegible] [illegible] [illegible] [illegible]

[illegible]

[illegible]

(2) INFORMATION FOR SEQ ID NO: [illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: linear

CTTCGCCGAA ACCTGATGCC GAGGAACAGG GTGTTCCCGT GAGCCCGACG GCGTCCGACC 60

CCGCGCTCCT CGCCGAGATC AGGCAGTCGC TTGATGCGAC AAAAGGGTTG ACCAGCGTGC 120

ACGTAGCGGT CCGAACAACC GGGAAAGTCG ACAGCTTGCT GGGTATTACC AGTGCCGATG 180

TCGACGTCCG GGCCAATCCC CTCCCGGCAA AGGGCGTATG CACCTACAAC GACGAGCAGG 240

GTGTCCCGTT TCGGGTACAA GCCGACAACA TCTCGGTGAA ACTGTTCGAC GACTGGAGCA 300

ATCTCAGCTC GATTTCTGAA CTGTCAACTT CACGCGTCCT CGATCCTGCC GCTGGCGTGA 360

TCGAGCTGCT GTCCGGTCTC ACGAACCTCC AAGCGCAAGG TACCGAAGTG ATACAAGGAA 420

TTTCGTCCAC CAAAATCACC GGGACGATTC GCGCGAGTTC TGTCAAGATG TTTGATCCTG 480

GCTTCAAGAG TGCAAGGCGT GCGACCGTGT GGATTGCTCA GGACGGCTCG TACCACCTCG 540

TCGAACGAG CATCGATCTC GGATCCGGCT CGATTCAGCT CACGCAGTCG AAATGGAACG 600

AACCCATCAA CGTCGACTAG GCCGAAGTTG CGTCGACGCG TTTCTCGAAA AGCCCTTGTG 660

AACGGTGTCA ACGGCATCCG AAAACTGACC CCCTGACGGC ATCTGAAAAT TGACCCCCTA 720

GACGCGGCGG TTGGTGGTTA TTCTTCGGTG GTTCCGCTCG GTGGGACGCG GCCGAGGTCG 780

CGGTCTTTGA GCCGGTAGCT GTCGCCTTTG AGGGCGACGA CTTCAGCATG GTGGACGAGG 840

CGGTCGATCA TGCCGGCAGC AACGACGTCG TCGCCGCCGA AAACCTCGCC CCACCGGCCG 900

AAGCGCTTAT TGCACGTGAC GATCAAGCTG CCCCGCTCAT ACCGGGAGGA CACCAGCTGG 960

AACAACAGGT TCGTGCGCTA CTCGTTAAAA CGAATGTAAC CGACTTCGTC AACCACCAGG 1020

ATTGATAGT CGCAAACTC GATGATTTAG TTTAGATGC CAGCGCTTC TTGACGCTCG 1080

CAAACTGCT CTACTATTA CCTCGCCCC TGACACCA TTCAACAAC GCCTGACAC 1140

[illegible] 1200

[illegible] 1260

[illegible] 1320

[illegible] 1380

[illegible] 1440

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GAATTCGGCA CGAGCCGGCG ATAGCTTCTG GGCCGCGGCC GACCAGATGG CTCGAGGGTT 60

CGTGCTCGGG GCCACCGCCG GGCGCACCAC CCTGACCGGT GAGGGCCTGC AACACGCCGA 120

CGGTCACTCG TTGCTTGTGG ACGCCACCAA CCCGGCGGTG GTTGCCTACG ACCCGGCCTT 180

CGCCTACGAA ATCGGTACA TCGNTGAAAG CGGACTGGCC AGGATGTGCG GGCAGAACCC 240

GGAGAACATC TTCTTCTACA TCACCGTCTA CAACGAGCCG TACGTGCAGC CGCCGGAGCC 300

GGAGAACTTC GATCCGGAGG GCGTGCTGGG GGGTATCTAC CGNTATCACG CGGCCACCGA 360

GCAACGCACC AACAAGGNGC AGATCCTGGC CTCCGGGGTA CCGATGCCCG CGGCGCTGCG 420

GGCAGCACAG ATGCTGGACG CGGAGTGGGA TGTCGCCGCC GACGTGTGGT CGGTGACCAG 480

TTGGGGCGAG CTAAACCGCG ACGGCGTGGT CATCGAGACC GAGAAGCTCC GCCACCCCGA 540

TCGGCCGGCG GGCGTGCCCT ACGTGACGAG AAGCCTGCAG AATGCTCGGG GCCCGGTGAT 600

CGCGGTGTCG GACTGGATGC GCGCGGTCCC CGAGCAGATC CGACCGTGGG TGCCGGGCAC 660

ATACCTCACC TTGGCACCCG ACGGCTTCGG TTTTTCCGAC ACTCGGCCGT CCGTTCGTCG 720

TTACTTCAAC ACCGACGCCG AACTCCAGTC TCGTCGCGAT TTTCGGGAGG GTTTGCCGGG 780

TCGACGGGTG AATACGGACC CATTCGGTCT CGGTTGCTGG CCGCCGCCGC AGTTACCGGG 840

ATCTGAGAAG CTTACTAGTA TTATTTTTCC TAAGTT 876

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

CAGATTCATA ACGAATTCAC AGCGGCACAA CAATATGTCG CGATCGCGGT TTATTTCGAC 120

AGCGAAGACC TGCCGCAGTT GGCGAAGCAT TTTTACAGCC AAGCGGTCGA GGAACGAAAC 180

CATGCAATGA TGCTCGTGCA ACACCTGCTC GACCGCGACC TTCGTGTCGA AATTCCCGGC 240

GTAGACACGG TGCGAAACCA GTTCGACAGA CCCCGCGAGG CACTGGCGCT GGCGCTCGAT 300

CAGGAACGCA CAGTCACCGA CCAGGTCGGT CGGCTGACAG CGGTGGCCCG CGACGAGGGC 360

GATTTCGTCG GCGAGCAGTT CATGCAGTGG TTCTTCGCAC AACAGATCGA AGAGCTGGCC 420

TTGATGGCAA CCCTGGTGCG GGTTGCCGAT CGGGCCGGGG CCAACCTGTT CGAGCTAGAG 480

AACTTCGTGG CACGTGAAGT GGATGTGCGG CCGGCCGCAT CAGGCGCCCC GCACGCTGCC 540

GCGGGCGCGC TCTACATCCC TGGGGCGGAT CAGCGAGTGG TCCCGTTCGC CGGCCGGTCT 600

TCCAGTCAGG CGTTCGGTGC GGCGGGGTGG TGAGTACCAA TCCAGGCCAC CCCGACCTCC 660

CGGNAAAAGT CGATGTCCTC GTACTCATCG ACGTTCCAGG AGTACACCGC CCGGCCCTGA 720

GCTGCCGAGC GGTCAACGAG TTGCGGATAT TCCTTTAACG CAGGCAGTGA GGGTCCCACG 780

GCGGTTGGCC CGACCGCCGT GGCCGCACTG CTGGTCAGGT ATCGGGGGGT CTTGCCGAGC 840

AACAACGTCG GCAGGAGGGG TGGAGCCCGC CGGATCCGCA GACCGGGGGG GCGAAAACGA 900

CATCAACACC GCACGGGATC GATCTCCGGA CGGCGCTGCG GGAATACCGA ACCGGTGTAG 960

GAGCGCCAGC AGTTGTTTTC CCACCAGCGA AGCGTTTTCG GTTCATCGGN GGCGNTTAAG 1020

T 1021

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

TCCGCCCCTG ANGTCCCGAC CGCCGCCCAG TGGACCAGNC TGTTCAACAG NCTCGNCGAT 240

CCCAACGTGT CGTTTGNGAA CAAGGGNAGT CTGGTCGAGG GNGGNATCGG NGGNANCGAG 300

GGNGNGNATC GNCGANCACA A 321

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 373 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TCTTATCGGT TCCGGTTGGC GACGGGTTTT GGGNGCGGGT GGTTAACCCG CTCGGCCAGC 60

CGATCGACGG GCGCGGAGAC GTCGACTCCG ATACTCGGCG CGCGCTGGAG CTCCAGGCGC 120

CCTCGGTGGT GNACCGGCAA GCGGTGAAGG AGCCGTTGNA GACCGGGATC AAGGCGATTG 180

ACGCGATGAC CGNGATCGGC CGCGGGCAGC GCCAGCTGAT CATCGGGGAC CGCAAGACCG 240

GCAAAAACCG CGGTCTGTGT CGGACACCAT CCTCAAACCA GCGGGAAGAA CTGGGAGTCC 300

CGTTGATCGC AAGAAGCAAG TGCGCTTGTC TATACGTTGG TCATCGGGCA AGAAGGGGAA 360

CTTACCACCG CGG 373

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible]
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GCGTGGAGGT TTTCGTCACC GTCAGCCGTG GNAAGTGGGA CACGTTGCGC GCCATNGNGT 300

TTGACGACGA NCCATATCGG NGATTCCCNC ACATNCGAAG TTCCGANGGA GA 352

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 776 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GAAATCCGCG TTCATTCCGT TCGACCAGCG GCTGGCGATA ATCGACGAAG TGATCAAGCC 60

GCGGTTCCGC GCGCTCATGG GTCACAGCGA GTAATCACCA AGTTCTCTGG TATATCGCAC 120

CTAGCGTTCA GTTGCTTGCC AGATCGGTTT CGTACCGTCA TCGCATGTAC CGGTTCGCGT 180

GCGGCACCCT CATCGTGGCC GCTTCCATCC TCGCCACGGG TGTGGCGGGT CTCGGGGTCG 240

GCGCGCAGTC CGCAGCCCAA ACTGCGCCGG TGCCCGACTA CTACTGGTGC CCGGGGCAGC 300

CTTTCGACCC CGCATGGGGG CCCAACTGGG ATCCCTACAC CTGCCATGAC GACTTCCACC 360

CCGACAAGGA CGGCTCCGAC GACAACGGCA ACTACCTCGG ACTCATCTTC GAAGGTCCCG 420

[illegible] 480

[illegible] 540

[illegible] 600

[illegible] 660

[illegible] 720

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CGCGACGACG ACGAACGTCG GGCCCACCAC CGCCTATGCG TTGATGCAGG CGACCGGGAT 60

GGTCGCCGAC CATATCCAAG CATGCTGGGT GCCCACTGAG CGACCTTTTG ACCAGCCGGG 120

CTGCCCGATG GCGGCCCGGT GAAGTCATTG CGCCGGGGCT TGTGCACCTG ATGAACCCGA 180

ATAGGGAACA ATAGGGGTGT GATTTGGCAG TTCAATGTCG GGTATGGCTG GAAATCCAAT 240

GGCGGGGCAT GCTCGGCGCC GACCAGGCTC GCGCAGGCGG GCGAGCCCGA ATCTGGAGGG 300

AGCACTCAAT GGCGGCGATG AAGCCCCGGA CCGGCGACGG TTCTTTGGAA GCAACTAAGG 360

AGGGCGCGGG CATTGTGATT CGAGTACCAC TTGAGGGTGG CGGTCGCCTG GTCGTCGAGC 420

TGACACCCGA CGAAGCCGCC GCACTGGGTG ACGAACTCAA AGGCGTTACT AGCTAAGACC 480

AGCCCAACGG CGAATGGTCG GCGTTACGCG CACACCTTCC GGTAGATGTC CAGTGTCTGC 540

TCGGCGATGT ATGCCCAGGA GAACTCTTCG ATACAGCGCT 580

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 160 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

[illegible]

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible]
(B) TYPE: [illegible]
(C) [illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GACACCGATA CGATGGTGAT GTACGCCAAC GTTGTCGACA CGCTCGAGGC GTTCACGATC 60

CAGCGCACAC CCGACGGCGT GACCATCGGC GATGCGGCCC CGTTCGCGGA GGCGGCTGCC 120

AAGGCGATGG GAATCGACAA GCTGCGGGTA ATTCATACCG GAATGGACCC CGTCGTCGCT 180

GAACGCGAAC AGTGGGACGA CGGCAACAAC ACGTTGGCGT TGGCGCCCGG TGTCGTTGTC 240

GCCTACGAGC GCAACGTACA GACCAACGCC CG 272

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 317 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCAGCCGGTG GTTCTCGGAC TATCTGCGCA CGGTGACGCA GCGCGACGTG CGCGAGCTGA 60

AGCGGATCGA GCAGACGGAT CGCCTGCCGC GGTTCATGCG CTACCTGCGC GCTATCACCG 120

CGTACGAGCT GAACGTGCGC GAAGCGCGCG GGTTCATCGT CTTCGACGTC GGGACGATCC 180

GTTCGGATCT CGCGTTGTTC GAGATGGTCT ATCTGCTACA TCGCTCTGCT CTCTGTCGT 240

GGAATCTGAC CGCGAAGATC AAGAAGCGGT CAAAGATCCA TGTGCTCGAC AGTCGCTTCG 300

CGCCCTGGTT CATCCGG 317

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 307 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

GCAGCGCCGG ACCACCTCGC CGGTGGTCAG CATGGTGATG ACCACGTCGG CGTCGGCGAC 120

CGCTTCGGGC GCGCTACGAA ACACCGCGAC ACCGTGCGCG GCGGCGCCGG ACGCCGCCGT 180

GG 182

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 308 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GATCGCGAAG TTTGGTGAGC AGGTGGTCGA CGCGAAAGTC TGGGCGCCTG CGAAGCGGGT 60

CGGCGTTCAC GAGGCGAAGA CACGCCTGTC CGAGCTGCTG CGGCTCGTCT ACGGCGGCCA 120

GAGGTTGAGA TTGCCCCGCG CGGCGAGCCG GTAGCAAAGC TTGTGCCGCT GCATCCTCAT 180

GAGACTCGGC GGTTAGGCAT TGACCATGGC GTGTACCCCG TGCCCGACGA TTTGGACGCT 240

CCGTTGTCAG ACCACGTGCT CGAACGCTTT CACCCGTGAA GCGCTACCTC ATCGACACCC 300

ACGTTTGG 308

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 367 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1539 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CTCGTGCCGA AAGAATGTGA GGGGACACGA TGAGCAATCA CACCTACCGA GTGATCGAGA 60

TCCTCGGGAC CTCGCCCCAC GGGGTCGACG CGGCAATCCA GGCCGCTCTG GCTCGAGCTG 120

CCTAGACTAT GCGCGCCCTG GACTGTTTCG AAGTACAGTC AATTCGAGGC CACCTGGTCG 180

ACGGAGCGGT CGCGCACTTC CAGGTGACTA TGAAAGTCGG CTTCCGCTGC AGCATTCCTG 240

AACCTTCAAG CGCGGCCGAT AACTGAGGTG CATCATTAAG CGACTTTTCC AGAACATCCT 300

GACGCGCTCG AAACGCCGTT CAGCCGACGG TGGCTCCGCC GAGGCGCTGC CTCCAAAATC 360

CCTGCGACAA TTCGTCGGCG GCGCCTACAA CGAAGTCGGT GCTGAATTCG TCGGGTATCT 420

GGTCGACCTG TGTGGGCTGC AGCCGGACGA ACCGGTGCTT GACGTCGGCT GCGGCTCGGG 480

GCCGATGGCG TTGCCGCTCA CCGGCTATCT GAACAGCGAG CGACGCTACG CCGGCTTCGA 540

TATCTCGCAC AAAGCCATCG CGTGGTCGCA CGAGTACATC ACTTCGCGCG CAACCGAACTT 600

CACGTTCGAG GTCTTCGACA TCAACAACCT CACTTACAAA CCGAAACCCA AAAACTACGC 660

ACTAGACTTT CGCTTTCCAT ACTACGATGC AGCCTTCGAT GTGGTGTTTC TCACCTCGGT 720

[illegible] 780

[illegible] 840

[illegible] 900

[illegible] 960

[illegible] 1020

[illegible] 1080

AGCCTGCGTG ATCGGTCATC ACCAACGGTG ACAGCAGCCG GTTGTGCACC ACCGCGAACG 1320

CCAGCCCGGT CTCCGGGTCT GTCCAGCCGA TCGAGCCGCC CAAGCCCACA TGACCAAACC 1380

CCGGCATCAC GTTGCCGATC GGCATACCGT GATAGCCAAG ATGAAAATTT AAGGGCACCA 1440

ATAGATTTCG ATCCGGCAGA ACTTGCCGTC GGTTGCGGGT CAGGCCCGTG ACCAGCTCCC 1500

GCGACAAGAA CCGTATGCCG TCGATCTCGC CTCGTGCCG 1539

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 851 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTGCAGGGTG GCGTGGATGA GCGTCACCGC GGGGCAGCCC GAGCTGACCG CCGCCCAGGT 60

CCGGGTTGCT GCGCCGGCCT ACGAGACGGC GTATCGCCTG ACGGTGCCCC CGCCGGTGAT 120

CGCCGAGAAC CGTCCTGAAC TGATGATTCT GATAGCGACC AACCTTTGCG GGCAAAACAC 180

CCCGGCGATC GACCTGAACG ACGCCGAATA CGGCAAGATG TGCCCCAAG ACGCCGCCGC 240

GATGTTTTGC TACCAGGTCG CGACCTCGAC GCTGAACGGC AAGTTCGTGC CGTTCGAGGA 300

GGCGCTGGAG ATGACAGTG CCTTGGCCT CTTCAAGAG CTGGTGCTG TCGAGGAGGC 360

CTTCGACACC GACCTGTTCA ACGAGCTGAT GAAGATGTTC CAGAAGCTG TGAAACAGTT 420

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 254 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GATCGATCGG GCGGAAATTT GGACCAGATT CGCCTCCGGC GATAACCCAA TCAATCGAAC 60

CTAGATTTAT TCCGTCCAGG GGCCCGAGTA ATGGCTCGCA CGACAGGAAC CTTACTGCTG 120

CGGGCACCTG TCGTAGGTCC TCGATACGGC CGAAGCCGTC GACATTTTCC ACCGACACCC 180

CCATCCAAAC GTTCGAGGGC CACTCCAGCT TGTGACCGAG CGACGCAGT CGCAGGCTGC 240

GCTTGGTCAA GATC 254

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1227 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CGCTGCTCAG CTTGGCGAAG GCGTGATCGG AGCGCTTGTC GCGCACGCCG TGGTCGATAC 600

CGCACAGCGC ATTGCGAACG ATGGTGTCCA CATCGCGGTT CTCCAGCGCG TTGAGGTATC 660

CCTGAATCGC GGTTTTGGCC GGTCCCTCCG AGAATGTGCC TGCCGTGTTG GCTCCGTTGG 720

TGCCGACCCC GTATATGATC GCCGCCGTCA TAGCCGACAC CAGCGCGAGG GCTACCACAA 780

TGCCGATCAG CAGCCGGTTG TGCCGTCGCT TCGGGTAGGA CAGCTGCCGC GGCACGCCGG 840

GATATGCGGC GGGCGGCAGC GCCGCGTCGT CTGCTCGTCC CGGGGCGAAG GCCGGTTCGG 900

CGGCGCCGAG GTCGTTGGGC TAGTCCAGGG CTTGGGGTTC GTGGGATGAG GGCTCGGCGT 960

ACGGCGCCGC TCGCTTTGTC TCCACAGCGT GTTTCGGCGA GTGGGACCTG GGCATTGTGG 1020

TTGTCCTAGG GTGGTGACGG TGACCAGCTG CTAGGGCGAC AATCGCCCGT CGCGTCAGCC 1080

GGCAGCATCG GCAATCAGGT GAGCTCCCTA GGCAGGCTAG CGCAACAGCT GCCGTCAGCT 1140

CTCAACGCGA CGGGGCAGGC CGCGGCGCCG ATAATGTTGA AAGACTAGGC AACCTTAGGA 1200

ACGAAGGACG GAGATTTTGT GACGATC 1227

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 181 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GCGGTGTCGG CGGATCCGGC GGGTGGTTGA ACGGCAACGG CGGTGTCGGC GGCCGGGGCG 60

GCGACGGCGT CTTTGCCGGT GCCGGCGGCC AGGGCGGCCT CGGTGGGCAG GGCGGCAATG 120

GCGGCGGCTC CACCGGCGGC AACGGCGGTC TTGGCGGCGC GGGCGGTGGC GGAGGCAACG 180

GCCCCGGACGG NGGCTTCGGT GGCAACGGCG GTAAGGGTGG CCAGGGCGGN ATTGGCGGCG 240

GCACTCAGAG CGCGACCGGC CTCGGNGGTG ACGGCGGTGA CGGCGGTGAC 290

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GATCCAGTGG CATGGNGGCT GTCAGTGGAA GCAT 34

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible] acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

[illegible]

(A) LENGTH: 53 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

ATGGCGTTCA CGGGGCGGCG GGGACCGGGC AGCCCGGNGG GGCCGGGGGG TGG 53

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 137 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GATCCACCGC CGGTGCAGAC GSTGCCCGTG GTGCCAGTTC GACCAGCCGC GGCAACGGCG 60

GCACCGGCGG CAACGCGCCC AACGCCACCC TYGTTGGNGC GGCCGGCGGC GCCGGCGGCA 120

AGGCTGGCAA CG 137

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 121 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 702 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CGGCACGAGG ATCGGTACCC CGCGGCATCG GCAGCTGCCG ATTCGICCGGG TTTCGCCACC 60

CGAGGAAAGC CGCTACCAGA TGGCGCTGCC GAAGTAGGGC GATCCGTTCG CGATGICCGGC 120

ATGAACGGGC GGCATCAAAT TAGTGCAGGA ACCTTTCAGT TTAGCGACGA TAATGGCTAT 180

AGCACTAAAG ACGATGATCT GATATGACGT ATCGGCAGAC CGTGACGGTG GATCAGCAAG 240

AGATTTTGAA CAGGGCCAAC GAGGTGGAGC CCCCGATGGC GGACCCACCG ACTGATGTCC 300

CCATCACACC GTGCGAACTC ACGGNGGNTA AAAACGCCGC CGAACAGNTG GTNTTGTCGG 360

CCGACAACAT GCGCCAATAC CTGGCGGCCG GTGCCAAAGA GCGGCAGCGT CTGGCGACCT 420

CGCTGCCCAA CGCGCCAAG GNGTATGCTG ATTTGATCA CGAGGCTGCG ACCGCCCTGG 480

ACAACGACGG CGAAGGAACT GTGCAGGCAC AACTGGCGCG GGCGTTCGGA GGGGACAGTT 540

CGGCCGAACT AACGGATACG CCGAGGGTGG CCACGGTCGG TGAACCGAAC TTCATGGATC 600

TCAAAGAAGC CGCAAGGAAG CTCGAAACCG CGGATAAGCT CGTATCGCTC CGCTACTTGN 660

TTGATCCTCT GAACATTNC ACTTGACTC CGAAGCAATA TC 702

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

CTGGGCGGCG GTGGCATGAG AATGCCGATG GGTGCTGCGC ATCAGGGACA AGGGGGCGCC 240

AAGTCCAAGG GTTCTCAGCA GGAAGACGAG GCGCTCTACA CCGAGGATCC TCGTGCCG 298

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1058 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CGGCACGAGG ATCGAATCGC GTCGCCGGGA GCACAGCGTC GCACTGCACC AGTGGAGGAG 60

GCATGACCTA CTCGCCGGGT AACCCCGGAT ACCCGCAAGC GCAGCCCGCA GGCTCCTACG 120

GAGGCGTCAC ACCCTCGTTC GCCCACGCCC ATGAGGGTGC GAGCAAGCTA CCGATGTACC 180

TGAACATCGC GCTGGCAGTG CTCGGTCTGG CTGCGTACTT CGCCAGCTTC GGCCCAATGT 240

TCACCCTCAG TACCGAACTC GGGGGGGGTG ATGGCGCAGT GTCCCGTGAC ACTGCCCTGC 300

CGGTCGGGGT GGCTCTGCTG GTTGCGCTCT TTGCCGGGGT GGTTCTGGTG CCTAAGGCCA 360

AGAGCATGT GACCGTAGTT [illegible] 420

[illegible] 480

[illegible] 540

[illegible] 600

[illegible] 660

[illegible] 720

[illegible] 780

[illegible] 840

[illegible] 900

[illegible] 960

[illegible] 1020

[illegible]

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 327 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TGGCACGAGA GACCGATGCC GCTACCCTCG CGCAGGAGCC AGGTAATTTC GAGCGGATCT 60

CCGGCGACCT GAAAACCCAG ATCGACCAGG TGGAGTCGAC GGCAGGTTCG TTGCAGGGCC 120

AGTGGCGCGG CGCGGCGGGG ACGGCGGCGC AGGTCGTCTT GGTGGCGTTC CAAGAAGCAG 180

CCAATAAGCA GAAGCAGGAA CTCGACGAGA TCTTGACGAA TATTCGTCAG GCCGGCGTCC 240

AATACTCGAG GGCCGACGAG GAGCAGCAGC AGGCGCTCTC CTCGCAAATG GGCTTCTGAC 300

CCGCTAATAC GAAAAGAAAC GGAGCAA 327

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 170 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GATCCGGCGG CACGGGGGGT GCCGGCGGCA GCACCGCTGG CGCTGGCGGC AACGGCGGGG 60

CCGGGGGTGG CGGCGGAACC GGTGGGTTGC TCTTCGGCAA CGGCGGTGCC GGCGGGCACG 120

GGGCCGT 127

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 81 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CGGCGGCAAC GGGGGCACCG CCGGCAACGG GAGCGGCTGC GCCGGCGGCA ACGGCGGCAA 60

CGGCGGCTCC GGCCTCAACG G 81

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 143 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CGGCACGAGA TCACACCTAC CGAGTGATCG AGATCGTCGG GACCTCGCCC GACGGTGTCG 60

ACCCCGNAAT CCAGGCCGGT CTGGCCCGAG CTGCTCAGAC CATGCCGCCC CTGGACTGGT 120

TCGAAGTACA GTCAATTCGA GGTACTTGG TCGACGGAGC GGTCGCCGCC TTCCAGGTGA 180

CTATGAAAGT CGGTTTCCGC CTGGAGGATT CCTGAACGTT CAAGCGCGGC CGATAACTGA 240

GGTGCATCAT TAAGCGACTT TTCTAGAACA TCCTGATGCA CTTGAAACGC GGTTCAGCCG 300

ACGGTGGCTG CGCCGAGGCC GTGCCTTCAA AATCCTGCCA ATAATTCGTC GCCGG 355

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 999 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GCCAACGGGG TGTCTGGAAG CGCGTCGTAT TACGAAGTCA AGTTCAGCGA TCCGAGTAAG 720

CCGAACGGCC AGATCTGGAC GGGCGTAATC GGCTCGCCCG CGGCGAACGC ACCGGACGCC 780

GGGCCCCCTC AGCGCTGGTT TGTGGTATGG CTCGGGACCG CCAACAACCC GGTGGACAAG 840

GGCGCGGCCA AGGCGCTGGC CGAATCGATC CGGCGTTTGG TCGCCCCCGC GCCGGCGCCG 900

GCACCGGCTC CTGCAGAGCC CGCTCCGGCG CCGGCGCCGG CCGGGGAAGT CGCTCCTACC 960

CCGACGACAC CGACACCGCA GCGGACCTTA CCGGGCCTGA 999

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 332 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Met His His His His His His Met His Gln Val Asp Pro Asn Leu Thr
1 5 10 15

Arg Arg Lys Gly Arg Leu Ala Ala Leu Ala Ile Ala Ala Met Ala Ser
20 25 30

Ala Ser Leu Val Thr Ala Asn Val Ile Ala Thr Val Asn Asp Asp Ile
35 40 45

Glu Pro Ala Pro Pro Val Ile Thr Thr Ala Ala Ser Ile Pro Ser Thr
50 55 60

[illegible]

[illegible]

[illegible]

145 150 155 160

Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg
165 170 175

Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala
180 185 190

Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro
195 200 205

Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val
210 215 220

Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys
225 230 235 240

Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn
245 250 255

Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly
260 265 270

Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu
275 280 285

Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro
290 295 300

Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Gly Gln Val Ala Pro Thr
305 310 315 320

Pro Thr Thr Pro Thr Pro Gln Asn Thr Leu Pro Ala
325 330

[illegible]

[illegible]

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala Ala Lys
1 5 10 15

Gly Arg

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val
1 5 10

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Gln Glu Ser Ile Ser Ile Xaa Xaa Xaa Leu Val Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

[illegible]

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

[illegible]

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: [illegible]
(C) STRANDEDNESS:
(D) TOPOLOGY: [illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 197 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Thr Gly Ser Leu Asn His Thr His Asn Arg Arg Ala Asn Gln Arg Lys
1 5 10 15

Asn Thr Thr Met Lys Met Val Tyr Ser Ile Ala Asn Gly Asn Thr Ala
20 25 30

Ala Ala Ala Leu Gly Ala Ala Asn Ala Gly Val Thr Ser Ile Met Ala
35 40 45

Gly Gly Pro Val Val Tyr Gln Met Gln Pro Val Val Phe Gly Ala Pro
50 55 60

Leu Pro Leu Asp Pro Ala Pro Asn Pro Asp Val Pro Thr Ala Ala Gln
65 70 75 80

Leu Thr Leu Leu Asn Ser Leu Ala Asp Pro Asn Val Ser Phe Ala
85 90 95

Asn Lys Gly Ser Leu Val Pro Ala Ala Thr Ala Gly Thr Ala Arg
100 105 110

Thr Asn Asp Asn Gly Leu Ile Ala Asn Ala Phe Gly Asp Leu Pro
115 120 125

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Asp Glu Val Ser Val Gly Thr Thr Ser Val Phe Asn Ala Asp Phe Leu
1 5 10 15

Ser Ile Leu Asp Ala Ile Asn Thr Ala Gly Thr Ile Ser Ala Val Ser
20 25 30

Gly Val Glu Gly Leu Ile Pro Gly Ser Asn Leu Leu Val Ile Lys Arg
35 40 45

Gly Ile Asn Ala Gly Ser Arg Ile Leu Leu Asp Ile Ala Thr Ser
50 55 60

Ala Gly Asn His Pro Asp Ser Asp Ile Phe Leu Asp Asn Val Thr Val
65 70 75 80

Ser Arg Asn His Ala Glu Ile Asn Leu Glu Asn Asn Glu Phe Asn Val
85 90 95

Val Asp Val Gly Ser Leu Asn Gly Thr Tyr Val Asn Asn Glu Pro Val
100 105 110

Asp Ser Ala Val Ile Leu Asn Asp Gly Asp Glu Ile Thr Phe Gly Gly Ser
115 120 125

Ile Leu Glu Ile Ile Gly Ile Ser Ser Thr Asp Asp Ile Ser
130 135 140

Ile Ile Gly Ile
145

Gly Pro Asp Arg Ser Ala Ser Leu Ser Leu Val Arg His Arg Arg Gln
25 30

Gln Arg Asp Ala Leu Cys Leu Ser Ser Thr Gln Ile Ser Arg Gln Ser
35 40 45

Asn Leu Pro Pro Ala Ala Gly Gly Ala Ala Asn Tyr Ser Arg Arg Asn
50 55 60

Ile Asp Val Arg Ile Ile Glu Met Leu Val Thr Ala Val Val Leu
65 70 75 80

Leu Val Cys Ser Gly Val Ala Thr Ala Ala Pro Lys Thr Tyr Cys Glu
85 90 95

Glu Leu Lys Gly Thr Asp Thr Gly His Ala Cys Gln Leu Gln Met Ser
100 105 110

Asp Pro Ala Val Asn Ile Asn Lys Ser Leu Ser Asp Tyr Tyr Lys Asp
115 120 125

Gln Lys Ser Leu His Asn Tyr Leu Ala Gln Thr Arg Asp Lys Phe Leu
130 135 140

Ser Ala Asn Thr Ser Ser Pro Val Glu Arg Val Ala Phe Tyr Glu Leu Asn
145 150 155 160

Ile Thr Ser Asn Ile Cys Thr Ala Glu Ile Ile Arg Gly Thr Gln
165 170 175

Val Val Lys Asn Thr Ala Arg Asp Ile Val Asp Ile Leu
180 185 190

Glu Glu Thr Ser Asp Ile Glu Ala Ile Asp Glu Leu Met Gln Ala Leu
195 200 205

Asn Lys Ser Ala Ala Pro Ala Arg Gln Asn Lys Val Lys Asp Leu Leu
210 215 220

Pro Gln Tyr Leu Ser Glu Val Asn Thr Leu Gln Thr Asn Val Thr Asp
225 230 235 240

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Thr Ala Ala Ser Asp Asn Phe Gln Leu Val Gln Gly Gly Gln Gly Phe
1 5 10 15

Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser
20 25 30

Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
35 40 45

Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val
50 55 60

Val Gly Ser Ala Pro Ala Asn Ser Leu Gly Ile Ser Thr Gly Asp Val
65 70 75 80

Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala
85 90 95

Asp Ala Leu Asn Gly Arg Ala Gly Ser Thr Val Asn Trp
100 105 110

Gln Thr Lys Ser Gly Gly Thr Arg Thr Val Asn Leu Ala Gln
115 120 125

Gly Pro Pro Ala
130

Ser Ala Gly Met Ala Arg Val Asn Asn Trp Xaa Val Met Pro Xaa Val
65 70 75 80

Ile Gln Ser Thr Xaa Ile Arg Xaa Xaa Gly Pro Phe Asp Asn Arg Gly
85 90 95

Ser Glu Arg Lys
100

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 103 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Met Thr Asp Asp Ile Leu Leu Ile Asp Thr Asp Glu Arg Val Arg Thr
1 5 10 15

Leu Thr Leu Asn Arg Pro Xaa Ser Arg Asn Ala Leu Ser Ala Ala Leu
20 25 30

Arg Asp Arg Leu Leu Ala Xaa Leu Xaa Asp Ala Glu Xaa Asp Asp Asp
35 40 45

Ile Asp Val Val Ile Leu Thr Gly Ala Gly Arg Ala Phe Val Ala Gly
50 55 60

Leu Asp Leu Leu Val Asp Gly Asp Arg Ala Ala Asp Arg Pro
65 70 75 80

Val Ser Leu Leu Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
85 90 95

Ala Ala Ala Ala Ala Ala Ala
100

(2) INFORMATION FOR SEQ ID NO:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 344 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:

215 220

Ala Ala Leu Ser His His Leu Asp Thr Ala Pro His Leu Pro Pro Pro
225 230 235 240

Thr Arg Gln Val Val Arg Arg Val Val Gly Ser Trp His Gly Glu Pro
245 250 255

Met Pro Met Ser Ser Arg Trp Thr Asn Glu His Thr Ala Glu Leu Pro
260 265 270

Ala Asp Leu His Ala [illegible] Thr Arg Leu Ala Leu Leu Thr Gly Leu Ala
275 280 285

Pro His Gln Val Thr Arg Arg Asp Val Ala Ala Ala Arg Ser Leu Leu
290 295 300

Asp Thr Asn Ala Ala Leu Val Gly Ala Leu Ala Trp Ala Ala Pro Thr
305 310 315 320

Ala Ala Arg Arg Thr Gly Thr Trp Leu Gly Ala Ala Ala Gly Gly
325 330 335

Val Pro Arg Gln Asn Pro Thr Gly
340

INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 495 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

Arg Glu Arg Tyr Leu Leu His Asp Gln Gln Trp Arg Ile Ala Glu Ser
85 90 95

Thr Gly Glu Leu Met Asp Arg Ser Ala Arg Cys Val Ala Ala Ala Glu
100 105 110

Asp Gln Tyr Glu Pro Gly Ser Ser Arg Arg Trp Ala Glu Arg Phe Ala
115 120 125

Thr Leu Leu Arg Asn Leu Gln Phe Leu Pro Asn Ser Pro Thr Leu Met
130 135 140

Asn Ser Gly Thr Asp Leu Gly Leu Leu Asn Gly Cys Phe Val Leu Pro
145 150 155 160

Ile Thr Asp Ser Leu His Ser Ile Ile Ala Thr Leu Gly Gln Ala Ala
165 170 175

Leu Leu Arg Asn Gly Trp Val Thr Gly Cys Asn Ile Ser His Leu
180 185 190

Arg Pro Ala Gly Asp Arg Val Ala Ser Thr Gly Gly Thr Ala Ser Gly
195 200 205

Ile Val Ser Ile Leu Arg Leu Tyr Arg Ser Ala Ala Gly Thr Val Ser
210 215 220

Met Gly Gly Arg Arg Arg Arg Ala Tyr Met Ala Val Leu Asp Val Ser
225 230 235 240

Ile Arg Asn Glu Tyr Asn Phe Ala Arg Ser Ile Ser
245 250

Ala Ile Val Asp Ala Gly Trp Gly Tyr Pro Asp Lys Glu Met Arg Asp
260 265 270

Ala Val Glu Leu Leu Asn Tyr Glu Lys Asn Lys Leu Val Arg Trp Leu
275 280 285

Tyr Ser Leu Pro Val Ser Leu Lys Pro Leu Phe Glu Cys Glu Val Gly
290 295 300

Lys Thr Leu Asn Ser Ser Glu Glu Pro Val Lys Lys Arg Ile Lys Asp
305 310 315 320

Leu Met Asp Ser Ala Met Thr Met Glu Glu Tyr Leu Lys Val Lys Glu
325 330 335

Leu Glu Asn Glu Leu Glu Ala Leu Glu Ala Val Leu Asp Glu Leu Asp
340 345 350

Phe Pro Leu Pro Leu Val Glu Phe
355 360

370 375 380

Asp Val Ile Asp Val Ser Arg Tyr Pro Phe Pro Glu Leu Gly Glu Ala
385 390 395 400

Ala Arg Ala Thr Arg Lys Ile Gly Leu Gly Val Met Gly Leu Ala Glu
405 410 415

Leu Leu Ala Ala Leu Gly Ile Pro Tyr Asp Ser Glu Glu Ala Val Arg
420 425 430

Leu Ala Thr Arg Leu Met Arg Arg Ile Gln Gln Ala Ala His Thr Ala
435 440 445

Ser Arg Arg Leu Ala [illegible] Arg Gly Ala Ile Pro Ala Phe Thr Asp
450 455 460

Ser Arg Pro Ala Arg Ser [illegible] Arg Arg Asp Ala Glu Val Thr Ser
465 470 475 480

Val Ala Pro Thr Gly
485

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

Thr Pro Thr Ala Ala Val Ser Pro Pro Val Leu Lys Gln Gly Asp
100 105 110

Asp Cys Pro Asp Ser Thr Leu Ala Val Lys Gly Leu Thr Asn Ala Pro
115 120 125

Gln Tyr Tyr Val Gly Asp Gln Pro Lys Phe Thr Met Val Val Thr Asn
130 135 140

Ile Gly Leu Val Ser Lys Lys Arg Asp Val Gly Ala Ala Val Leu Ala
145 150 155 160

Ala Tyr Val Tyr Ser Leu Asp Asn Lys Arg Leu Trp Ser Asn Leu Asp
165 170 175

Cys Ala Pro Ser Asn Gln Thr Leu Val Lys Thr Ser Ser Pro Gly Glu
180 185 190

Gln Val Thr Thr Ala Val Ser Trp Pro Gly Met Val Ser Ala Pro Arg
195 200 205

Ser Pro Leu Pro Arg Pro Ala Ile Gly Leu Gly Thr Tyr Asn Leu Val
210 215 220

Val Ala Leu Gly Asn Leu Arg Ser Leu Val Pro Phe Thr Leu Asn
225 230 235 240

Gln Pro Pro Pro Pro Ile Gly Leu Val Leu Ala Pro Gly Leu Ala Gln
245 250 255

Lys Val Asp Asp Arg Pr Ile Asn Ser Ala Asp Ala Leu Val Ala Ala
50 55 60

Val Arg Ser Lys Ala Pro Gly Ala Thr Val Ala Leu Thr Phe Gln Asp
65 70 75 80

Pro Ser Gly Gly Ser Arg Thr Val Gln Val Thr Leu Gly Lys Ala Glu
85 90 95

Gln

145 150 155 160

Gln Ile Gln Ala Leu Asn Ser Gly Thr Asp Leu Pro Ile Thr Pro Ile
165 170 175

Ser Val Ile Phe Arg Ser Asp Lys Ser Gly Thr Ser Asp Asn Phe Gln
180 185 190

Lys Tyr Leu Asp Gly Val Ser Asn Gly Ala Trp Gly Lys Gly Ala Ser
195 200 205

Glu Thr Phe Ser Gly Gly Val Gly Val Gly Ala Ser Gly Asn Asn Gly
210 215 220

Thr Ser Ala Leu Leu Gln Thr Thr Asp Gly Ser Ile Ile Tyr Asn Gln
225 230 235 240

Trp Ser Phe Arg Val Gly Leu Gln Leu Ala Ser Asn Gln Leu Ile Thr
245 250 255

Ser Ala Gly Pro Asp Pro Val Ala Ile Ile Thr Thr Gln Ser Val Gly Lys
260 265 270

Thr Ile Ala Gly Ala Lys Ile Met Gly Gln Gly Asn Asp Leu Val Leu
275 280 285

Asp Thr Ser Ser Ile Tyr Arg Phe Thr Ile Pro Gly Ser Tyr Pro Ile
290 295 300

Val Leu Ala Thr Tyr Val Ile Ser Val Ile Tyr Trp Asp Ala Ile
305 310 315 320

Thr Gly Tyr Ala Val Arg Gln Leu Met Ile Ala Ile Gly Leu Gly
325 330 335

Ile Gly Leu Arg Ala Ile Ser Ser Ile Leu Lys Asp Ile
340 345 350

Gln Ala Asn Arg Gly Arg Asn Val Arg Asn Thr Gly His Ala Glu Asp
1 5 10 15

Gln Thr His Gln Asp Arg Leu His His Gly Cys Arg Arg Ala Ala Val
20 25 30

Val Val Arg Gln Asp Arg Ala Ser Val Ser Ala Thr Ser Ala Arg Pro
35 40 45

Pro Arg Arg His Pro Ala Gln Gly His Arg Arg Arg Val Ala Pro Ser
50 55 60

Gly Gly Arg Arg Arg Pro His Leu His His Val Glu Pro Asp Asp Arg
65 70 75 80

Arg Asp Arg Pro Ala Leu Leu Asp Arg Thr Gln Leu Ala Glu His Pro
85 90 95

Asp Pro His Arg Arg Gly Leu Ala Asp Pro Gly Arg Val Arg Gly Arg
100 105 110

Gly Arg Leu Arg Arg Val Asp Arg Gly Arg Leu Gln Leu Asp Arg Asp
115 120 125

Ala Asp His Gly Ala Leu Thr Arg Gly Arg Gly Leu His Arg Gly Val
130 135 140

Gln His Arg Gly Gly Leu Val Leu Arg Arg Val Pro Gly Val Arg
145 150 155 160

Cys Arg His Arg Arg Gly Glu Arg Arg Val Ala Arg Pro Gly Gln Gly
165 170 175

Arg Val Leu Arg His Arg Arg Leu Arg Leu Val Ala
180 185 190

Ala Val Asp Arg Leu Arg Pro Gly Val Ser Arg Arg Asp Gly Arg Val
195 200 205

290 295

Asn Arg Pro Arg Arg
305

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 580 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

Met Ala Val Trp Cys Thr Asn Gly Ile Thr Gly Arg His Arg His Gly
1 5 10 15

Arg Cys Arg Val Arg Ala Ser Gly Thr Arg Ser Ser Asn Arg Trp Cys
20 25 30

Ser Thr Thr Ala Asp Tyr Leu Ala Ser Leu Thr Pro Thr Glu Ala Arg
35 40 45

Leu Pro Leu Thr Arg Arg Ile Thr Tyr Tyr Ser Ile Arg Val Gly Cys
50 55 60

Arg Ile Arg Leu Ile Phe Ser Arg Arg Leu Ala Leu Ala Arg Ile
65 70 75 80

Ser Arg Thr Ser Leu Thr Val Arg Ile Leu Ala Ser Ala Ala Ser
85 90 95

Ile Arg Ala Glu Ile Leu Leu Ala Ala Ser Ala Ala Ala Ala Ala
100 105 110

Gly Leu Ile Pro Gly Val Ile Pro Thr Met Thr Pro Pro Pro Gly Met
195 200 205

Val Arg Gln Arg Pro Arg Ala Gly Met Leu Ala Ile Gly Ala Val Thr
210 215 220

Ile Ala Val Val Ser Ala Gly Ile Gly Gly Ala Ala Ala Ser Leu Val
225 230 235 240

Gly Phe Asn Arg Ala Pro Ala Gly Pro Ser Gly Gly Pro Val Ala Ala
245 250 255

Ser Ala Ala Pro Ser Ile Ile Ala Ala Asn Met Pro Pro Gly Ser Val
260 265 270

Gln Gln Val Ala Ala Gly Val Val Pro Ser Val Val Met Leu Leu Thr
275 280 285

Asp Leu Gly Arg Ser Ser Gly Gly Gly Ile Ile Leu Ser Ala
290 295 300

Glu Gly Leu Ile Leu Thr Asn Asn His Val Ile Ala Ala Ala Asn Ala Lys
305 310 315 320

Pro Pro Leu Gly Ser Ser Ile Ile Thr Thr Val Thr Ile Ser Asp
325 330 335

Gly Arg Thr Ala Pro Phe Thr Val Val Gly Ala Asp Pro Thr Ser Asp
340 345 350

Leu Ala Val Leu Arg Val Gln Gly Val Ser Gly Leu Thr Pro Ile Ser
355 360 365

Leu Gly Ser Ser Ala Asp Leu Met Val Gly Glu Pro Val Ile Ala Ile
370 375 380

Gly Asn Pro Phe Gly Leu Gly Gly Thr Val Thr Thr Gly Ile Val Ser
385 390 395 400

Ala Leu Gly Arg Ser Gly Leu Asn Ile Glu Gly Leu Ile Gln Thr Asp
405 410 415

Val Gln Val Thr Asn Asp Lys Asp Thr Pro Gly Ala Lys Ile Val Glu
500 505 510

Val Val Ala Gly Gly Ala Ala Ala Asn Ala Gly Val Pro Lys Gly Val
515 520 525

Val Val Thr Lys Val Asp Asp Arg Pro Ile Asn Ser Ala Asp Ala Leu
530 535 540

Val Ala Ala Val Arg Ser Lys Ala Pro Gly Ala Thr Val Ala Leu Thr
545 550 555 560

Thr Gln Asp Ile Ser Gly Gly Ser Ala Ser Val Gln Val Thr Leu Gly
565 570 575

Leu Ala Glu Gln
580

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 233 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

Val Gln Gly Asp Asn Ile Ser Val Lys Leu Ile Asp Asp Trp Ser Asn
115 120 125

Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val Leu Asp Pro Ala
130 135 140

Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn Leu Gln Ala Gln
145 150 155 160

Lys Thr Glu Val Ile Asp Gly Ile Ser Thr Thr Lys Ile Thr Gly Thr
165 170 175

Leu Pro Ala Ser Ser Val Lys Ile Leu Asp Pro Gly Asn Lys Ser Ala
180 185 190

Arg Pro Ala Thr Val Trp Ile Ala Gln Asp Gly Ser His His Leu Val
195 200 205

Arg Ala Ser Ile Arg Leu Glu Asp Ser Ile Ile Thr Gln Ser
210 215 220

Lys Trp Asn Glu Pro Val Asn Val Asp
225 230

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

(ii) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Val Pro Pro Ala Pro Pr[illegible] Leu Pro Pr[illegible] Leu Pro Pro Ser Pro Ile Ser
1 5 10 15

Tyr Ala Ser Pr[illegible] Pro Ser Pro Pro [illegible] Leu Pro Tyr Ala Pro Pro Val Ala
20 25 30

Pro Gly Pro Pr[illegible] Met [illegible] Pro Leu Asp Pro Trp Pro Pro Ala Pro Pro
35 40 45

Leu Pro Tyr Ser Thr Pro Pro Gly Ala Pro Leu Pro Pro Ser Pro Pro
50 55 60

Ser Pro Pro Leu Pro
65

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

Gly Ile Val Ile Asp Ile Asn Gly Val Val Leu Thr Asn Asn His Val
85 90

Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln
100 105 110

Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val Ala
115 120 125

Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile Gly
130 135 140

Gly Gly Val Ala Val Gly Asp Pro Val Val Ala Met Gly Asn Ser Gly
145 150 155 160

Gly Gln Gly Gly Glu Pro Ser Tyr Arg Val Val Ala Leu
165 170

Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Gln Gln Thr
180 185 190

Leu Asn Gly Leu Ile Gln Thr Asp Ala Ala Ile Gln Pro Gly Asp Ser
195 200 205

Gly Gly Pro Val Val Asn Gly Ser Gln Val Val Gly Met Asn Thr
210 215 220

Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe Ala
225 230 235 240

Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Asp Gly
245 250 255

Val Val Asn Asn Ala Gln Arg Ala Val Leu Gly Val Gln Val
260 265 270

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 205 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Gly Asp Ser Ile Trp Ala Ala Ala Asp Gln Met Asn Asn Gly Phe Val
1 5 10 15

Leu Gly Ala Thr Asn Gly Arg Thr Thr Leu Thr Gly Val Gly Leu Gln
20 25 30

His Ala Asp Gly His Ser Leu Leu Leu Asp Ala Val Asn Pro Arg Val
35 40 45

Val Ala Tyr Asp Pro Ala Ile Ala Tyr Glu Ile Gly Tyr Ile Lys Glu
50 55 60

Ser Gly Leu Ala Asn Met Tyr Gly Val Asn Pro Glu Asn Ile Ile Phe
65 70 75 80

Tyr Ile Thr Val Tyr Asn Glu Pro Tyr Val Gln Pro Pro Gln Pro Glu
85 90 95

Asn Ile Asp Pro Val Gly Arg Leu Gly Val Tyr Arg Tyr His Ala
100 105 110

Ala Thr Ile Gln Asn Thr Asn Ile Lys Leu Ile Ala Thr Val
115 120 125

Ala Met Ser Ala Ala Ala Asp Asn Met Ala Asn Ile Ala
130 135 140

Ala Gln Ser Thr Val Gly Arg Gly Phe Gly Ala Gly Ile Pro Gly Arg
245 250 255

Arg Val Asn Ile Asp Pro Phe Gly Ala Gly Arg Gly Pro Pro Ala Gln
260 265 270

Leu Pro Gly Phe Asp Gln Gly Gly Gly Leu Arg Pro Xaa Lys
275 280 285

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

[illegible]

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

Arg Asn Ala Leu Gln Leu Gln Ala Pro Ser Val Val Xaa Asp Gln Gly
35 40 45

Val Lys Glu Pro Leu Xaa Thr Gly Ile Lys Ala Ile Asp Ala Met Thr
50 55 60

Pro Ile Gly Arg Gly Gln Arg Glu Leu Ile Ile Gly Asp Arg Lys Thr
65 70 75 80

Gly Lys Thr Ala Ile Ala Ile Asp Thr Ile Ile Asn Gln Arg Glu
85 90 95

Ile Leu Cys Ile Tyr Val Ala Ile Gly Gln Lys Glu Ser Thr Val Arg
100 105 110

Gln Val Val Glu Thr Leu Arg Lys His Gly Ala Met Glu Tyr Thr Ile
115 120 125

Val Val Ala Ala Thr Ala Ser Asp Ala Ala Pro Leu Gln Tyr Leu Ala
130 135 140

Pro Tyr Ser Gly Cys Ser Met Gly Glu Tyr Phe Arg Asp Asn Gly Lys
145 150 155 160

His Ala Leu Ile Ile Tyr Asp Asp Leu Ser Lys Gln Ala Val Ala Tyr
165 170 175

Arg Gln Met Ser Leu Leu Leu Arg Arg Pro Pro Gly Arg Glu Ala Tyr
180 185 190

Pro Gly Asp Val Phe Tyr Leu His Ser Arg Leu Leu Glu Arg Ala Ala
195 200 205

Lys Leu Asn Asp Ala Leu Gly Gly Gly Ser Leu Thr Ala Leu Pro Val
210 215 220

Ile Glu Thr Gln Ala Gly Asp Val Ser Ala Tyr Ile Pro Thr Asn Val
225 230 235 240

Ile Ser Ile Thr Asp Gly Gln Ile Phe Leu Glu Thr Glu Leu Phe Tyr
245 250 255

Lys Gly Ile Arg Pro Ala Ile Asn Val Gly Leu Ser Val Ser Arg Val
260 265 270

Gly Ser Ala Ala Gln Thr Arg Ala Met Lys Gln Val Ala Gly Thr Leu
275 280 285

Lys Leu Glu Leu Ala Gln Tyr Arg Glu Leu Glu Ala Phe Ala Gln Phe
290 295 300

Ala Ser Asp Leu Asp Ala Ala Thr Gln Asn Gln Leu Val Arg Gly Gln
305 310 315 320

Arg Leu Thr Glu Leu Leu Lys Gln Gly Gln Tyr Ser Pro Met Ala Ile
325 330 335

Glu Glu Gln Val Ala Val Ile Tyr Ala Gly Thr Arg Gly Tyr Leu Asp
340 345 350

Ala Leu Glu Val Gly Asp Val Leu Arg Phe Glu Ser Ala Phe Leu Ser
355 360 365

His Val Val Ser Gln His Ala Asp Leu Leu Asp Thr Ile Arg Thr Thr
370 375 380

Gly Asp Leu Ser Asp Glu Ile Glu Ala Lys Leu Lys Glu Ala Leu Thr
385 390 395 400

Asn Phe Lys Ala Thr Phe Ala
405

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) [illegible]

(xi) [illegible]

Arg Ile Val Arg Gly
115

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 103 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Met Tyr Arg Ile Asn Tyr Arg Thr Leu Met Leu Asn Asn Tyr Ile Leu
1 5 10 15

Ala Thr Thr Tyr Thr Ala Gly Leu Gly Val Gly Ala Leu Ser Asn Ala
20 25 30

Thr Ala Pro Val Pro Asp Tyr Tyr Trp Tyr Ile Gly Ile Ile Leu Asp
35 40 45

Ile Ala Thr Gly Pro Asn Ile Asp Pro Tyr Ile Tyr Ile Asp Arg Phe
50 55 60

His Asn Leu Ser Leu Leu Leu Ile Ala Thr Ile Ser Ala Asn Ile Gly Pro
65 70 75 80

Ile Ile Ala Leu Ile Leu Ala Gly Thr Leu Ala Ala Ile Ile Asp
85 90 95

Ala Asp Glu Ala Arg Asn Gly Gly Ile Ala Arg Ile Trp Arg Glu His
30

Ser Met Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala
35 40 45

Thr Lys Glu Gly Arg Gly Ile Val Met Arg Val Pro Leu Glu Gly Gly
50 55 60

Gly Arg Leu Val Val Gly Leu Thr Pro Asp Glu Ala Ala Ala Leu Gly
65 70 75 80

Arg Thr Leu Leu Lys Gly Val Thr Pro
85

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 95 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Thr Asp Ala Arg Thr Leu Ala Glu Thr Ala Gly Arg Leu Arg Ile
1 5 10 15

Tyr Asp Ala Leu Gly Ala Glu Arg Ala Ala Thr Ala Arg Ala Gly
20 25 30

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
35 40 45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn
1 5 10 15

Arg Ala Asn Glu Val Gln Ala Pro Met Ala Asp Gln Pro Thr Asp Val
20 25 30

Pro Ile Thr Pro Tyr Gln Ile Thr Lys Asn Lys Asn Ala Ala Glu Glu
35 40 45

Met Val Leu Ile Ala Asp Ala Ala Ser Ser Ile Ala Ala Gly Ala
50 55 60

Leu His Arg Ala Arg Leu Ala Ile Leu Ala Arg Asn Ala Ala Leu Xaa
65 70 75 80

Lys Gly Glu Val Asp Gln Glu Ala Ala Thr Ala Leu Ala Asn Asp Gly
85 90 95

Ser Gly Thr Val Gln Gly Ser Ala Val Gly Asp Ser
100 105 110

Val Ala Glu Ile Thr Asp Asp Pro Ile Val Ala Thr Ala Gly Glu Ile
115 120 125

Val Leu Met Asp Ile Val Ser Val Asp Asp Ala Asp
130 135 140

Ile Val Ala Ala Ile Ile Ala Ala Ile Ala Ile Thr
145 150 155

Thr Leu Leu Asp Gly Asp
160

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 263 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Ser Ser Gly Leu Gly Gly Gly Val Ala Ala Asn Leu Gly Arg Ala Ala
225 230 235 240

Ser Val Arg Tyr Gly His Arg Asp Gly Gly Lys Tyr Ala Xaa Ser Gly
245 250 255

Arg Arg Asn Gly Gly Pro Ala
260

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 408 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Met Thr Tyr Ser Ile Ala Asn Ile Ala Tyr Ser Ile Ala Glu Ile Ala
1 5 10 15

Gly Ser Tyr Arg Gly Val Ile Ile Ser Ile Ala His Ala Asp Asn Gly
20 25 30

Ala Ile Ala Ile Ala Ile Ala Ile Ala Ile Ala Ile Ala Ile Asn Gly
35 40 45

Leu Ala Ala Ala Ile Ala Leu Ala Ile Ala Met Ile Ser Ile Gly
50 55 60

Ala Ile Ala Ile Ala Ile Ala Ile Ala Ile Ala Ile Ala Ile Ala Ile
65 70 75 80

Ala Pro Ala Pro Asn Pro Lys Ile Asn Pro Tyr Gly Gln Tyr Gly Arg
165 170 175

Tyr Gly Gln Tyr Gly Gln Tyr Gly Val Gln Pro Gly Gly Tyr Tyr Gly
180 185 190

Gln Gln Gly Ala Gln Gln Ala Ala Gly Leu Gln Ser Pro Gly Pro Gln
195 200 205

Gln Ser Ile Gln Pro Pro Gly Tyr Gly Ser Gln Tyr Gly Gly Tyr Ser
210 215 220

Ser Ser Ile Ser Gln Pro Gly Ser Tyr Tyr Thr Ala Gln Pro Pro Ala
225 230 235 240

Gln Pro Pro Ala Gln Ser Gly Ser Gln Ala Ser His Gln Gly Pro Ser
245 250 255

Ala Ile Pro Thr Gly Phe Ile Ser Thr Ser Pro Pro Pro Pro Val Ser
260 265 270

Asn Gly Thr Gly Ser Gln Ala Gly Ser Ala Ile Val Asn Tyr Ser Asn
275 280 285

Pro Ser Gly Gly Gln Gln Ser Ser Ser Pro Gly Gly Ala Pro Val
290 295 300

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible]
(B) TYPE: [illegible]
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

[illegible] 480

GAGTTGCTGC AGGCCTAGCT GAACTGA 507

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 168 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

[illegible]

(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:[illegible]:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible] 300

[illegible] 360

[illegible] 420

[illegible] 480

[illegible] 500

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible] acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

Glu Asn Ala Ala Met Ala Ala Gln Leu Lys Ala Val Thr Gly Ala Ala
65 70 75 80

Gln Tyr Ile Gly Leu Val Glu Ser Val Ala Gly Ser Cys Asn Asn Tyr
85 90 95

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 756 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

ATGAAAAAAT ACCTTCTAAA TTTTGCTTTT ATTGCTTCTT TTGTAAGCGA AATTTAGGGA 60

AATCTCAGTT CTATTTATTC TTCTTTTTAT TATTTAAATC ATTCCTTGAT TAAGTTCGCA 120

TTTTTTTTTT CCTTAGCCT TTCTGAAATC TA 152

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 51 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible] acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible] 300

(2) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible] acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

[illegible]

[illegible]

TGATCGCCGA GAACCGTGCT GAACTGATGA TTCTGATAGC GACCAACCTC TTGGGGCAAA 780

ACACCCCGGC GATCGCGGTC AACGAGGCCG AATACGGCGA GATGTGGGCC CAAGACGCCG 840

CCGCGATGTT TGGCTACGCC GCGGCGACGG CGACGGCGAC GGCGACGTTG CTGGTCGTTG 900

AGGAGGCGGC GGACATGACC ATCGCGCTGC TTATTTTGTA CGTAGCCGCT TTCGTAGGG 960

[illegible] 1020

[illegible] 1080

[illegible] 1140

[illegible] 1200

[illegible] 1260

[illegible] 1320

[illegible] 1380

[illegible] 1440

[illegible] 1500

[illegible] 1560

[illegible] 1620

[illegible]

GCCGGCAACA ACATGGCGCA AATCGACAGT GTCAATGGCT TGAGTGGGAT CTGACAGCAA 2340

GCCAAGGCCA GGGACGTGGT TACGAGTGA AGTTCGTCGC GTGATCGTTC GGGTGGCAGT 2400

CTAAGTGGTC AGTGCTGGGG TGTTGGTGGT TTGCTGCTTG GCGGGTTCTT CGGTGCTGGT 2460

CAGTGCTGCT CGGGCTCGGG TGAGGACGTC GAGGCCCAGG TAGCGCCGTC CTTCGATCCA 2520

TTCGTCGTGT TGTTCGCCGA AGACGGCTCC GACGAGGCGG ATGATCGAGG CGTGGTCGGG 2580

GAAGATGCCC ACGACGT GG TTCGCGTCG TAACTCTCGC TTGAGGGGTT CCTCGGGGTT 2640

CTTGACGTAG ATTTGCCTGT AGATCTCCTT GGAAGGCG GTGAACGCCA TCAGGTCGGT 2700

GAGGCGGTTG TCGACCTCGT CGACGTTGAT GGGATTTTG TCGCTCAGAG CGTCGAGTAC 2760

CTGATCATAT TCGCAACAA CTGATTCGCC CTCCTCCTGT TCGTAGATGT ACTTCAGCAG 2820

GTTCTGGCAC CACGGCGA GGGCTTCGGC GTCGTGCTCA ATGACGATTG TCTTGAGTG 2880

GTTTTTGCAG CACTGCCAGG CGGCTCGCCG CACCTGACC CGGATTGCGG CTACGAGGGC 2940

CCGCTGGCGC TCGCTGCTCA CAAGCCGAAC CTTGACGAGC TCGCGGGCGA CGACGTCGCC 3000

GAGGAAGC ACGCAGTCGC CCTCGTTCTC GTAGAGCTTG ACTTGGATCA CTACGATC 3058

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible]
(B) TYPE: [illegible]
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

Ala Gly Gln Ala Glu Leu Thr Ala Ala Gln Val Asn Val Ala Ala Ala
85 90 95

Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala
100 105 110

Glu Asn Arg Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly
115 120 125

Gln Asn Thr Pro Ala Ile Ala Val Asn Gln Ala Glu Tyr Gly Glu Met
130 135 140

Trp Ala Gln Asp Ala Ala Ala Ala Gly Thr Gly Tyr Ala Ala Ala Thr Ala
145 150 155 160

Thr Ala Thr Ala Thr Leu Leu Phe Gly Ala Thr Met Thr
165 170 175

Ser Ala Gly Gly Leu Leu Gln Ala Ala Val Gln Gly Ala Ser
180 185 190

Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu
195 200 205

Gln Gln Leu Ala Gln Gln Thr Ala Trp Gln Ile Pro Gly Leu Lys Leu
210 215 220

Gly Gly Leu Trp Phe Ile Val Ala Gln Arg Leu Ile Ile Asn
225 230 235 240

Val Val Ser Met Ala Leu Ala Leu Met Leu Met Leu Thr Asn Ser Gly Val
245 250 255

Ala Met Thr Ala Leu Ile Val Ala Met Leu Gly Phe Val Ile Ala
260 265 270

Xaa Xaa Leu Leu Gly Val Leu Ala Xaa Val Ile Ile Ala Pro Xaa Tyr Val Met
370 375 380

Pro His Ser Ile Ala Ala Gly
385 390

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1725 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

GAGGTAATA CTTCGGTTT AGGGTTGGA CTTGATCTGT TTCAAGGTG 60

AAGTCCTTG CGGTTTGCT GCCGTGGAT TAGATTAAT TGTTTTTT ATAGAACTA 120

ATTTTTTA ATTTTTGTT AAGTATAAA TTTATTATT GTTAATTA CTTTCAAG 180

TTTTTTTT TTAATTTT TTTATTAT TTTATAAT TTTTAGTG TGTTAGGG 240

TTAATTAA TTTTTAAT ATTAATAA TTAATTAT TTTTTTAT TTTTTTAA 300

ATAAAAAA AAATTTTT TTTAATT TAATTTA ATTTTT TTTTTTA 360

TATTTT TTTTTTTT TTTTTTTT TTTTTTTA TTTTTTAT 420

TATTTTTT TTTTTTTT TTTTTTTT TTTTTTTT TTTTTTTT 480

TATTTTTT TTTTTTTT TTTTTTTT TTTTTTTT TTTTTTTT 540

[illegible] 1140

GGCGTTGCTG CGGTTCGAGC ACGCCCACCT GATTACCAAC CCCGGCGGCC TCCTTGAGCA 1200

GGCCGTCGCG GTCGACGAGC CCATCGACAC CGCCCCGGCG AACCAGTTGA TGAACAATGT 1260

GCCCCAAGCG CTGCAACAGC TGGCCCAGCC AGCGCAGGGC GTCGTACCTT CTTCCAAGCT 1320

GGCAGGGCTG TGGACGTCGT TCTACGCGCA CTGTTCGCCG CTGAGCAACG TCAGTTCGAT 1380

ATCCAACTAC CACATCTTCA TGATGGACA [illegible] ATGACGAACA TCTTGCACTC 1440

[illegible] 1500

[illegible] 1560

[illegible] 1620

[illegible] 1680

CAGCTTGACC AGCCGCTGCC AAACCGCTTT TGATACATC CTGGG 1725

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

[illegible]

(B) TYPE: [illegible]
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

[illegible] 60
[illegible] 120
[illegible] 180
[illegible] 240
[illegible] 300
[illegible] 360
[illegible] 420
[illegible] 480
[illegible] 540
[illegible] 600
[illegible] 660
[illegible] 720
[illegible] 780
[illegible] 840

AGGACAGCTC CAATCTGTCT ATCGCGGCTA TCATCGTGAT GTCGATGTTA TCAAAGGTTG 1440

ACGGCAATTG GCTGATCACC AAGTTCACCC CGGTTTAGGT TGTCCGTAGGC GGTCGCCAAG 1500

TCTGACGGGG GCGCGGGTGG CTGCTCGTGC GAGATACCGG CCGTTCTCCG GACAATCACG 1560

GCCCGACCTC AAACAGATCT CGGCGGCTGT CTAATCGGCC GGGTTATTTA AGATTAGTTG 1620

CCACTGTATT TATCTGATGT TCAGATTCTT TATTGGATTT TAGCTTCGCC GCAGGGCGGC 1680

TGGTGCACTT TGTATCTTTC GTTTTTACTA TTTGACAGAA TTTGATCTTT TCCCGACCTT 1740

GTTTCTTCTC CATCATCTGT TTATTTACTT ATTGAGCGGA AGGATTATTC AAGTGGTGGA 1800

TTTTGGCCCC TTTCTCATCT TTTTAACTTT TTTAAGAATA TACCCTGTTT CGGGTTCGGC 1860

TTCGTTGTTA GCATTCTCTA ATTTTCTTTA CATTTTCTTT ATTCATTTGT TTTTGGCCCC 1920

GTCGGATTTT CAGTCGGTCT TTGGGGTTTT GACTACTTTA TTTTTATAAC TTTTTTCGGT 1980

GGGTCTGATT GTGGCGGCGG CTTTGCTCTA TTCTTTCTCT ATCATCGTCA CTGTCGGGGA 2040

GCTATGATTA ACCGCTGTAT AGCTTTGGTT TGTTTTCGCC GCCTACGAGA CGGCGTATGC 2100

GTTCATTTTA TTTTTTTTTT TTTTTTTTTT AATTTTTTTT GAACTGATTA TTCTGATAGC 2160

TATTAACTTT TTGGGTCAAA ATATTTTTTT TATTTTTTTC AATCATTTTG AATACGTGGA 2220

CATTTGTTTT TAATATTTAT TTTTACTTTT TTTTTTTTTT TTTACGTTTT TGATCGTTGC 2280

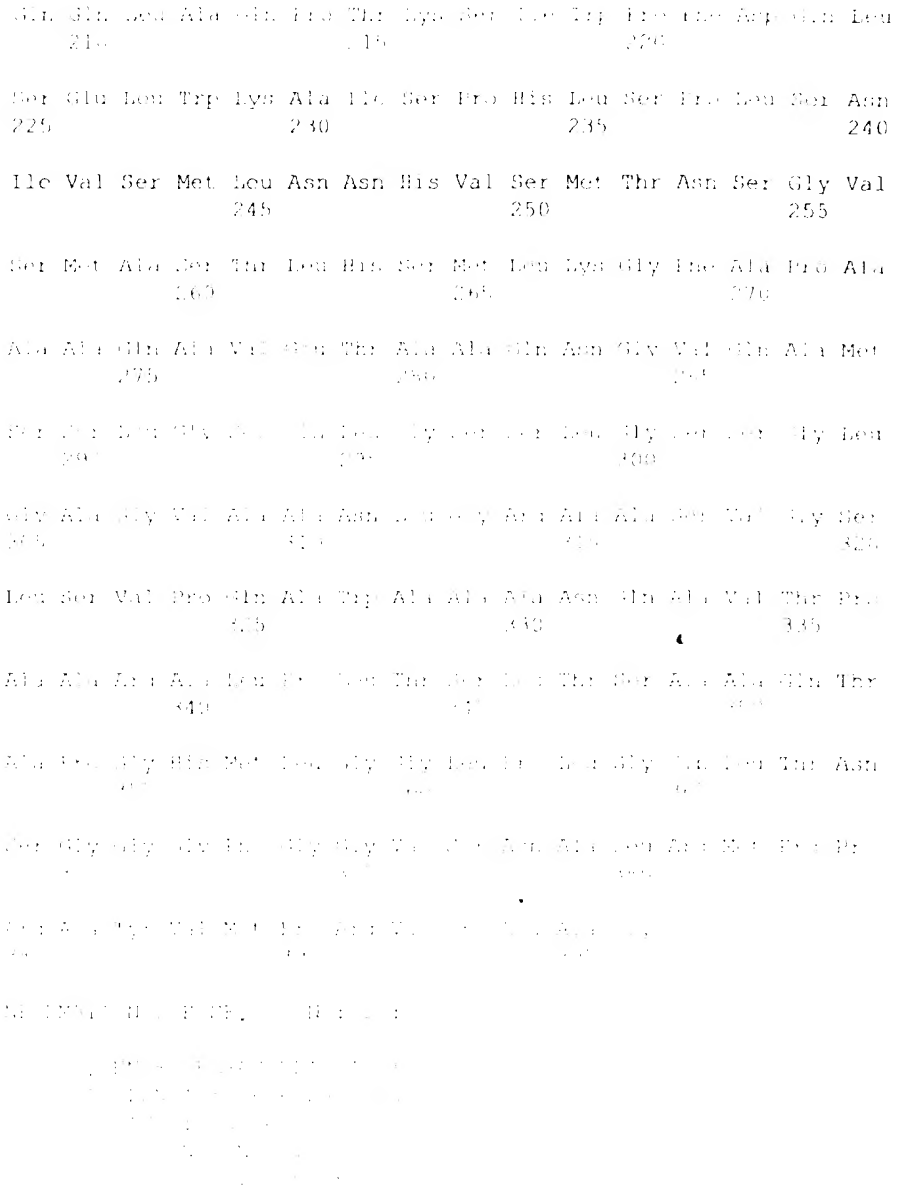
TGATTTTTTT TTTTTTATTA ATATTTTTTA TTTTTTTTAT GATTTTTTTG TTTTTTTTTT 2340

TTTTTTTTAA TTTTTTATTA TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 2400

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 2460

(2) INFORMATION FOR SEQ ID NO:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 396 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear



AGCCTGGACT GGAGTTGCTA [illegible] GTTGAGGTT GTAAGGCGGA TTTGGTCTG 240

GCTACAAACC GCGTCAACAC AGGCAAGAGC [illegible] CACGGACGGG CGCAAGCCGC 300

GGCATACACC CAGGCCATGG CCACGACGCC GTCGCTGCCG GAGATCGCCG CCAACCACAT 360

CACCCAGGCC GTCCTTACGG CCACCAACTT CTTCGGTATC AACACGATCC CGATCGCGTT 420

GACCGAGATG GATTATTTCA TCGTTATGTG [illegible] GAACTACGCA GCCCTGGCAA TGGAGGTCTA 480

[illegible] ACCTCGTTTA ACACGTTTTT [illegible] GAGCGATGCT [illegible] GATCCT 540

TGGTCGCGAT GCGACTCAAA [illegible] GTTGATTTC GAATGCCCT [illegible] GCAG 600

[illegible] GTTTGGAATT [illegible] [illegible] TTACTTAG ACGCCGGTC AACTTGGTGA 660

GATGAGTGGC GGATTGATCT AGCTCATTAA [illegible] GAGGTGACCT [illegible] TTTCAG 720

[illegible] CGCACCGCTC GCCAAAATT [illegible] AAAGTGGCT [illegible] 780

[illegible] AATCGCTTT CGAATTATTC [illegible] GATCAGGTC [illegible] 840

[illegible] GTGCGTCGCG AGTTGCTTCT [illegible] TGGTGGTTGA GCGGCACGCC 900

[illegible] [illegible] AAAAAT [illegible] [illegible] GCATTGCCGT CGATTGCTGC 960

[illegible] [illegible] [illegible] GGAGCGATGG GCCAGGGTGC 1020

[illegible] [illegible] [illegible] 1080

[illegible] [illegible] [illegible] 1140

[illegible] [illegible] [illegible] 1200

[illegible] [illegible] [illegible] 1260

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

CTAGTGGATG GGACCATGGC CATTTTCTGC AGTCTCACTG GTTCTGTGGT TGACATTTTG 60

GTACGCTGGT GTAAATGAAT CACTGGTTCT AAAAGGGTTG TGCTTGGGTA TATCGTTGGG 120

AATTTCTATA GTTCTTTGTG GTCGTAAGAG TTTTCTGTAT TCGGCCGGCA TGACAACGTC 180

TGAATTTTTT TTAAATTTTA TAAAATGATA AAAGTTGATT ACTCATGTAA AGTCAAACTC 240

GTCCATTTTT GTGTTTCTTT ATTCTTATTT AATTTCTGGT TTCTCTATTC AAACATCTCA 300

TGATTTTTAT TTTTTTTTA ATTGATTTTT TTTTGATTTT TTTGATTTTT TGTCTTGAGG 360

TTCTTCAGGT TTTGTTTAAT TTGTGTATGT TTGTTTTTTT ATTTTTTTTT GATACTTTGT 420

TATTCTTGGG AA 432

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

Thr Gln Ala Met Ala Thr Thr Ile Ser Leu Pro Gln Ile Ala Ala Asn
100 105 110

His Ile Thr Gln Ala Val Leu Thr Ala Thr Asn Phe Phe Gly Ile Asn
115 120 125

Thr Ile Pro Ile Ala Leu Thr Glu Met Asp Tyr Phe Ile Arg Met Trp
130 135 140

Asn Gln Ala Ala Leu Ala Met Glu Val Tyr Gln Ala Gly Thr Ala Val
145 150 155 160

Asn Thr Leu Thr Thr Tyr Leu Gln Pro Met Ala Ser Ile Leu Asp Pro
165 170 175

Gly Ala Ser Ile Ile Asp Thr Thr Asn Ile Ile Pro Gly Met Ile Ser
180 185 190

Gly Ser Ser Thr Pro Val Gly Asn Leu Ile Ala Ala Thr Thr Thr
195 200 205

Leu Gly Gln Leu Gly Gln Met Ser Gly Ile Met Gln His Leu Thr Gln
210 215 220

Ile Leu Asn Val Thr Ser Leu Ile Asn Gln Thr Gly Thr Gly
225 230 235 240

Gly Asn Ile Ala Asp Ala Ile Gln Ala Ala Ile Gly Leu Gly
245 250 255

Thr Ile Ile Ala Arg Ile Leu Ala Gly Leu Gly Ile Asn
260 265 270

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly
275 280 285

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 100 amino acids
(b) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Ala Glu Met Lys Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly
1 5 10 15

Asn Glu Leu Phe Lys Lys Asp Gly Asp Asp Lys Thr Gln Leu Asp Gln Val
20 25 30

Leu Ser Thr Ala Gly Ser Leu Gln Leu Val Trp Arg Gly Ala Ala Gly
35 40 45

Thr Ala Ala Gln Ala Ala Val Val Asn Phe Gln Glu Ala Ala Asn Lys
50 55 60

Gln Lys Gln Glu Leu Asp Glu Leu Ser Thr Asn Ile Arg Gln Ala Gly
65 70 75 80

Val Gln Tyr Ser Arg Ala Asp Glu Glu Gln Gln Ala Leu Ser Ser
85 90 95

Gln Met Gly Phe
100

TGACCCGCT AATACGAAAA GAAACGAAT AAAAATTCA CACATATTA CCAAATTTT 300

ACGCTATCG AGCCCGCGC AAGCGCAAT CACGAAAATC TTACTTCAT TATTCTTC 360

CTTGACGAGG GCAAGCAGTC CCTGACCAAG CTCGCA 396

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

Ile Ser Gly Ala Leu Asp Glu Thr Thr Ile Arg Glu Val Glu Ser Thr Ala
1 5 10 15

Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln
20 25 30

Ala Ala Val Val Arg Ile Asn Glu Glu Ala Ala Asp Gln Glu Ser Gln
35 40 45

Leu Asp Gln Ile Ile Thr Asn Leu Arg Ala Ala Gly Val Glu Tyr Ser
50 55 60

(2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

CCTGGTACTC CTCTCAACCT CTACGCCACC AGGCTCTCCC GAGCTTCTTC AGGGACTGCT 240

TCCCCTCGTC AAGGAGGGAA TGAATGGACG TGACTATTTC CTGGATTGCG CTTGCCGCGG 300

CCTCGATACC CGCGAAATTC CACTGCTGCT CTGTCATGTT TTTGCTCCGT TTCTTTTCGT 360

ATTAGCGGGT CAGAAGCCCA TTTGCGA 387

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 372 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

TGGTACGACG ATTCCGTTTA TTTAACCCCT CTTCCTTAGT GCTCTGTTTC GTCCGTGAGC 60

TGCGGGTCCA ATTTTCTTCT TTTTCCCA TTTCCTTGGA TTGATGGACT ACTTTCTACT 120

TTCGCGACCT TTCTTTCCGT CTCTTTCTGA TAGCCTTCAT CCCTGCCGCT ACTTGCGTAC 180

TCTTTTTTCT TGGCTTCTTC CTTCCTTTCC TTTTTTGACC CAGATTTTCT TGATTAACCC 240

CTTTTTTCCT TCCTTGATTC TATTTTTTCA 270

(2) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1773 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser
1 5 10 15

2. INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

Ala Ala Met Leu Ala Ala Leu Gly Val Gly Leu Ala Ala Leu
1 5 10

2. INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

(A) LENGTH: 14 [illegible] acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

Asp Ile Gly Asn Gln Asn Thr Gln Asp Gln Glu Xaa Ala Val
1 5 10

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Ala Xaa Ala Asn Leu Asn Ile Xaa Thr Xaa Ile Xaa Val Ile
1 5 10

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Ala Pro Lys Thr Tyr Xaa Glu Glu Leu Lys Gly Thr Asp Thr Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Glu Leu Thr Ser
1 5 10 15

Leu Leu Asn Ala Leu Leu Ala Ala Gly Val Leu Asp Ala Leu
20 25 30

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

(C) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Asp Pro Gly Tyr Thr Pro Gly
1 5

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) FEATURE:
(A) OTHER INFORMATION: /note= "The Xaa at location 1 can be either a
Pro or Thr"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Xaa Xaa Gly Tyr Thr Pro Gly Xaa Xaa Xaa
1 5 10

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

Xaa Xaa Xaa His Lys Pro Phe Leu Arg

(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Xaa Asp Xaa Xaa Lys Xaa Ala Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:

(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Ala Pro Glu Ser Gly Ala Gly Leu Gly Gly Thr Val Gln Ala Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Xaa Tyr Ile Ala Tyr Xaa Thr Thr Ala Gly Ile Val Ile Gly Lys Ile
1 5 10 15

Asn Val His Leu Val
20

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

GTCCTGGTAC CTCATCGGCT ATGGTGTGGA TCCCGGCTTC GCCTCGCCTC GAAGCGCCGG 300

CCGTGCCCGG TGTTGTGCCT GCCCCGGTGC CAATCCCGGT CCCGATCATC ATTCGCCCGT 360

TCCCGGGTTG GCAGCCTGGA ATGCCGACCA TCCCCACCGC ACCGCCGACG ACGCCGGTGA 420

CCACGTCGGC GACGACGCCG CCGACCACCC CGCCGACCAC GCCGGTGACC ACGCCGCCAA 480

CGACGCCGCC GACGACGTCG GTGACCACGC CGCCAACGAC GCCGTCGACT ACGCCGGTGA 540

CCACGCCACC AACGACCGTC GCCCCGACGA CCCTCCGCCC GACGACGTCG TTCCGACCA 600

CGTTGCCCTC GACGACGGTC GCTCCAGCCA CGCCAACGCC GACCACCGTC GTTCCGCAGC 660

CGCCCTACGT CCTCGACTCAA CAACAACCCT GACCCTGACC AACCTAGTAC TACTACGTGG 720

CCCCGCAGAC GGTGGCGCCG GCTCCGTACC CGCATCCTCC TGCCCGCAAC CCTAGCGGCG 780

CCGCCGACTT ATTAGCGCCG TTCCTGATGA CCTCGCCGCT TCATTACGCC CTAGCGACAT 840

GGCCGGTGAT GCGGTGACCG TGGTGGTGCC CTGTTCTAAC GA 882

(2) INFORMATION FOR SEQ ID NO:184:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 815 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

AGCGGCTAGC TGTCGGTTTT CGTCAAGCTT GATCATGATG TCGAGGTGAC TTTACCGCG 600

CCCCCGAAG GAGGCGCTGA ACTCGGCGTT GAGCCGATCG GCGATCGGTT GGGGCAGTGC 660

CCAGGCCAAT ACGGGGATAC CGGGTGTCNA AGCCGCCCCG AGCGCAGCTT CGGTTGCGCG 720

ACNGTGGTCG GGGTGGCCTG TTACGCCGTT GTCNTCGAAC ACGAGTAGCA GGTCTGCTCC 780

GGCGAGGGCA TCGACCACGC GTTGGCTCAC CTCGT 815

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1152 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

ACGAGCGGCC GTTGAGGTTC TCAGATCAAA GATCTTTGGA ACTTACGGTT GCTTTCAGC 60

CTTCTCCCAG AACAACTGCC GAAGTCTCTC GTCGTCAAA CAGCGCTTCA TTTACCTCC 120

TATCACGGGT TCAATGACCA GATCATCGCC GACCTTGATA TCGTACGCGT CGGTTACACA 180

TTTCAAAGCA CTTCTCAGCC TATCGCTTCC GATCTCGGCC TTCTGACGCA TTTCGTCGAC 240

CTTCTTCGCA TCTTCTTCAC TGAGAATGCC GTCAAAGCCA ACGTCTTCGA CAACGTTCG 300

GTCGTCCTCA CCTTTTTTCG CGTCGCCGCA ATCAAATCCT TTCATCATC TCCTCCTTAC 360

[illegible] 420

[illegible] 480

[illegible] 540

[illegible] 600

GCTGAAACTG CTGGCTGCCT CTGAGTCCTC GTACGTAAAT CCTCTGCCAA CAGGCCGGCC 960

CGCCGGCCGA ATACCAGCAG ATCGGACAGC GAATTGCCGC CCAGCCGGTT GGAGCCGTGC 1020

ATACCGCCGG CACACTCACC GGCAGCGAAC AGGCCTGGCA CCGTGGCGGC GCCGGTGTCC 1080

GCGTCTACTT CGACACCGCC CATCACGTAG TGACACGTCC GCCCGACTTC CATTGCCTGC 1140

GTTCGGCACG AG 1152

(2) INFORMATION FOR SEQ ID NO:136:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 956 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

ATTCTCGAGA TTGGAAGGG TCTACTTCCT GCTTCTTCAA GCCGCATTAC TCTCGAACGAC 60

CAGCAATCGA CGAAGAGCA GATCCTCTTT GATCAGTCT ACGCGGTCTA CTCGGCGCAT 120

CCCCTCGAGT CCGCCTCCTC CGCCTTTTC TTCAAC TCAACGAAC CTTTTCCCCC 180

CGTCTTTAC CCCCCATCT CCCCCCCCC CCCCCCCCC CCATCC CCAACCTCC 240

ATCCAAGAC CCAAGAAC CTCTCTCCA ACTTTCCT CCCCCCC CCAACCTT 300

CCCCCAACC CCACCACCC GACTCTCCC CAACCCC CTCACCCC CCCCTCCC 360

CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC 420

CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC 480

CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC 540

CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC 600

CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC 660

CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC 720

CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC 780

CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC 840

CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC 900

CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCCCCC CCCCCC 956

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Asn Ala Val Val Ala Phe Ala Val Ile Gly Phe Ala Ser Leu Ala Val
1 5 10 15

Ala Val Ala Val Thr Ile Arg Pro Thr Ala Ala Ser Lys Pro Val Glu
20 25 30

Gly His Gln Asn Ala Gln Pro Gly Lys Phe Met Pro Leu Leu Pro Thr
35 40 45

Gln Gln Gln Ala Pro Val Pro Pro Pro Pro Pro Asp Asp Ile Thr Ala
50 55 60

Gly Phe Gln Gly Gly Thr Ile Pro Ala Val Gln Asn Val Val Pro Arg
65 70 75 80

Pro Gly Thr Ser Pro Gly Val Gly Gly Thr Pro Ala Ser Pro Ala Pro
85 90 95

Leu Ala Pro Ala Val Ile Gly Val Val Ala Pro Val Ile Pro
100 105 110

Val Pro Ile Pro Ile Ile Ile Pro Gly Trp Gln Pro Gly Met Pro
115 120 125

Pro Ile Pro Thr Ala Pro Ile Ile Ser Val Thr Ile Ala Thr
130 135 140

Ser Pro Pro Thr Ile Pro Thr Pro Ser Ala Thr
145 150 155

Gly Ser Gly Gly Gly Asp Leu Phe Gly Gly Phe
260 265

(2) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 174 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

Ile Asn Gln Pro Leu Ala Pro Lys Ala Pro Pro Asp Pro Pro Ser Pro
1 5 10 15

Pro Arg Pro Pro Val Pro Pro Val Pro Pro Leu Pro Pro Ser Pro Pro
20 25 30

Ser Pro Pro Thr Gly Trp Val Ser Ala Ala Leu Leu Pro Pro Trp Leu
35 40 45

Ala Gly Thr Ser Pro Ala Pro Ala Val Pro Met Ala Pro Leu Pro
50 55 60

Pro Ala Ala [illegible]
65 70 75 80

Ser His [illegible]
85

Ala [illegible]
100

[illegible]
115

[illegible]
130

[illegible]
145 150 155 160

[illegible]
165 170

(A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

Xaa Pro Pro Ala Gly Val Ser Asp Thr Arg Val Ser Gly Leu Thr Gly
1 5 10 15

Ala Val Gln Pro Ser Pro Arg Thr Thr Ala Glu Asp Pro Arg Pro Arg
20 25 30

Arg Arg Arg
35

(2) INFORMATION FOR SEQ ID NO:140:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 104 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Xaa Asp Asp Xaa Xaa Xaa Xaa Xaa Arg Trp Xaa Xaa Xaa Xaa Xaa

106

(2) INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(v) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

GAATCGATAT GGGCCATCAT CATCATCATC ACGTGATCGA CATCATCTTC ACC 53

(2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(v) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

GGATCCTGCA GGCTCGAAAC CACCGAGCGG T 31

(2) INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(v) FRAGMENT TYPE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

CTCGAATTC AGCGTTGGAA ATTCTGGTCA C 31

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(iv) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "PCR primer"

(v) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

GAGAGAATTC TCAGAAGCCC ATTTGGAGT A A 33

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1000 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(v) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1000

(x) PUBLICATION INFORMATION:

TTT CAC GAG AGG TAT CCG AAC GTC ACG ATC ACC GCT CAG GGC ACC GGT 412
Phe His Glu Arg Tyr Pro Asn Val Thr Ile Thr Ala Gln Gly Thr Gly
75 80 85

TCT GGT GCC GGG ATC GCG CAG GCC GCC GCC GGG ACG GTC AAC ATT GGG 460
Ser Gly Ala Gly Ile Ala Gln Ala Ala Ala Gly Thr Val Asn Ile Gly
90 95 100

GCC TCC GAC GCC TAT CTG TCG GAA GGT GAT ATG GCC GCG CAC AAG GGC 508
Ala Ser Asp Ala Tyr Leu Ser Glu Gly Asp Met Ala Ala His Lys Gly
105 110 115

CTG ATG AAC ATC GCG CTA GCC ATC TCG GCT CAG CAG GTT AAC TAC AAC 556
Leu Met Asn Ile Ala Leu Ala Ile Ser Ala Gln Gln Val Asn Tyr Asn
120 125 130 135

CTG CCT GGA GTG AGC CAG CAC CTT AAG TTG AAT GGA AAA GTT CTG GCC 604
Leu Pro Gly Val Ser Gln His Leu Lys Leu Asn Gly Lys Val Leu Ala
140 145 150

GCC ATG TAC CAG GGC ACC ATC AAA ACC TGG GAT GAT CCG CAG ATC GCT 652
Ala Met Tyr Gln Gly Thr Ile Lys Thr Trp Asp Asp Pro Gln Ile Ala
155 160 165

GCG CTG AAC [illegible] [illegible] GTG AAC [illegible] [illegible] [illegible] ACC [illegible] GTA GTT [illegible] CTG 700
Ala Leu Asn [illegible] Gly Val Asn [illegible] [illegible] Gly Thr Ala Val Val Pro Leu
170 175 180

CAC [illegible] [illegible] GAC [illegible] [illegible] GGC GAT [illegible] [illegible] [illegible] [illegible] ACC [illegible] TAC CTG 748
His Arg Ser Asp [illegible] Ser Gly Asp Thr [illegible] [illegible] Phe Thr Gln Tyr Leu
185 190 195

[illegible] 796
[illegible] Tyr [illegible] Asn [illegible]
200 205 210

[illegible]

280 290

GAG GCA ATT TCA ATG ATC GAC GGA CCA GCA CCG GAC GGC TAC CCA ATC 1084
Glu Ala Ile Ser Met Ile Asp Gly Pro Ala Pro Asp Gly Tyr Pro Ile
300 305 310

ATC AAC TAC GAG TAC GCC ATC GTC AAC AAC CGG CAA AAG GAC GCC GCC 1132
Ile Asn Tyr Glu Tyr Ala Ile Val Asn Asn Arg Gln Lys Asp Ala Ala
315 320 325

ACC GCA CAG ACT TCA CAG GCA TTT [illegible] CAA TGG GCA ATC ACC GAC GGC 1180
Thr Ala Gln Thr Ser Gln Ala Phe [illegible] Gln Trp Ala Ile Thr Asp Gly
330 335 340

AAC AAG GCT TCA TTC TAC GAC GTT [illegible] TTT CAA CCG CTG CCC CCA 1228
Asn Lys Ala Ser Phe Tyr Asp Val [illegible] Phe Gln Pro Leu Pro Pro
345 350 355

GCA GTC GTA AAA TCC TCA GAT GCA CCA [illegible] ATA TCA ATT TC AGC 1276
Ala Val Val Lys Ser Ser Asp Ala Pro [illegible] Ile Ser Ile Ser Ser
360 365 370

TAG TGTTT GAAAAAAA AGAAGAAA TTTATTGGG TAATGGGGT GTTTGGGGA 1333

GAATGTTGC [illegible] GAA [illegible] TGGTT GGGG 1384

CACAACCC [illegible] TTGTTTTG [illegible] TTGGTGGG CTGGTGATC 1434

AAAAAT [illegible] TAAA [illegible] TTTTTT [illegible] 1484

AAA AAAA [illegible] AA [illegible] TTTTTT [illegible] 1534

TACT [illegible] 1584

[illegible] 1634

[illegible] 1684

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

Val Lys Ile Arg Leu His Thr Leu Leu Ala Val Leu Thr Ala Ala Pro
1 5 10 15

Leu Leu Leu Ala Ala Ala Gly Cys Gly Ser Lys Pro Pro Ser Gly Ser
20 25 30

Pro Glu Thr Gly Ala Glu Asn Gly Thr Val Ala Thr Thr Pro Ala Ser
35 40 45

Ser [illegible] Val Thr Leu Ala [illegible] Thr [illegible] Thr Leu Leu Ser Ile Leu
50 55 60

Ile Ala Leu Tyr [illegible] Ile Ala Ile [illegible] Arg Lys Ile Asn Val Thr
65 70 75 80

Ile Thr Ala [illegible] Gly Thr [illegible] Ser [illegible] Lys Leu Ala Ala Ala
85 90 95

Ala Gly Thr Val Asn Ile Gly Ala Ser Asn Ala Tyr Leu Ser Ala Gly
100 105 110

Asp Met Ala Ala Ile Lys Gly Met Met Asn Leu Ala Leu Asn Ile Ser
115 120 125

Asn Ile Ile Val Asn Tyr Asn Leu [illegible] Val [illegible] Ile Lys
130 135 140

Leu Asn Lys Ile Val Leu Ala Asn Met [illegible] Gly Thr [illegible]
145 150 155

Thr Asn Asn Pro [illegible] Ile Asn Leu [illegible]
160 165

Phe Leu Leu Pro Asn Asn Gln Ser Leu Gln Ala Ala Asn Ala Gly Phe
275 280 285

Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
290 295 300

Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn
305 310 315 320

Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
325 330 335

Met Thr Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp Gln Val
340 345 350

Lys Phe Gln Arg Leu Phe Pro Asn Val Val Lys Leu Ser Asp Ala Leu
355 360 365

Leu Ala Thr Ile Leu Leu
370

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1403 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

TGACACCTTC TTGTTCACCC AGTACGTGTC CAAGCAAGAT CCCGAGGGCT GGGGCAAGTC 780

GCCCGGCTTC GGCACCACCG TCGACTTCCC GGCGGTGCCC GGTGCGCTGG GTGAGAACGG 840

CAACGGCGGC ATGGTGACCG GTTGCGCCGA GACACCGGGC TGCGTGGCCT ATATCGGCAT 900

[illegible] 960

[illegible] 1020

[illegible] 1080

[illegible] 1140

[illegible] 1200

[illegible] 1260

[illegible] 1320

[illegible] 1380

[illegible] 1440

[illegible] 1500

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

Met Lys Ile Arg Leu His Thr Leu Leu Ala Val Leu Thr Ala Ala Pro
1 5 10 15

Leu Leu Leu Ala Ala Ala Gly Gly Gly Ser Lys Pro Pro Ser Gly Ser
20 25 30

Pro Gln Ile Gly Ala Gly Ala Gly Ser Val Ala Thr Thr Pro Ala Ser
35 40 45

Ser Pro Val Thr Leu Ala Thr Thr Gly Ser Thr Leu Ile Tyr Pro Leu
50 55 60

Pro Arg Leu Lys Lys Ile Ala Ile Ser Glu Arg Tyr Ile Asn Val Thr
65 70 75 80

Ile Pro Ala Glu Gly Thr Gly Ser Gly Ala Gly Ile Ala Glu Ala Ala
85 90 95

Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Ile Ser Gln Gly
100 105 110

Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala Ile Ser
115 120 125

Ala Gln Val Asn Asn Thr Arg Ile Ile Val Ala Gln Ala Lys
130 135 140

Val Asn Ile Tyr Thr Ile Ala Gly Tyr Ala Lys Thr Gly Thr Thr Lys
145 150 155 160

Thr Arg

Ser Gln Arg Gly Leu Gly Ile Ala Gln Leu Gly Asn Ser Pro Gly Asn
260 265 270

Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe
275 280 285

Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
290 295 300

Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Gln Tyr Ala Ile Val Asn
305 310 315 320

Asn Arg Gln Lys Asp Ala Ala Ile Ala Gln Thr Leu Gln Ala Phe Leu
325 330 335

Arg Ala Ile Thr Arg Gly Asn Lys Ala Ser Phe Leu Asp Gln Val
340 345 350

His His Gln Pro Leu Pro Pro Arg Val Val Lys Leu Pro Asp Ala Leu
355 360 365

Ile Ala Thr Ile Ser Ser
370

(2) INFORMATION FOR SEQ ID NO:169:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1777 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

TGCGATCTGA TCGGGATGGG CGCGTCGGAC AAGCTCAGCC CATCGGGACC CGACGGCTAT 660

AGCTATGGCG AGCAACGAGA CTTTTTGTTC GCGCTCTGGG ATGCGCTCGA CCTCGGCGAC 720

CACGTGGTAC TGGTGCTGCA CGACTGGGGC TCGGCGCTCG GCTTCGACTG GGCTAACCAG 780

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

GACATTGAAT CGTACCGCTC TCCTTAGCGG CTCCTCCGCT TGAATGCCTA TATCACGCAC 60

GCCCATGTTC TGGCTGTCGA CCTTCGCCCC ATGCCCGGAC GTTGGTAAAC CCAGGGTTTG 120

ATCAGTAATT CCGGGGGACG GTTGCGGGAA GCCGGCCAGG ATGTGCGTGA GCCGCGGCGC 180

CGCCGTCGCC CAGGCCACCG CTGGATGCTC ATCCTCGGTT CGGCGACGTA GCCAGGCGTT T 240

GCTCCTCTCT CTCTACATCG GTACTCCGCT GATACGCCGC CGGCATTCAT GCCTCAAGAC 300

CCTGAACAAC CCCCCCAATT CACA 324

(2) INFORMATION FOR SEQ ID NO:154:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1348 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

[illegible] 900

CGGTTAACCG GCCAGAGGGT CTGGACGCCA TCGAGGATCT GCACATCTGG ACCGCCGAGT 960

CGGTGCGCGC CGACCGGCTC GACTTTCGGC CCAAGCACAA ACTGGCCGTC TTGGTGGTCT 1020

CGGCGATCCC GCTGGCCGAG CCGGTCCGGC TGGCGCGTAG GCCCGAGTAC GGCGGTTGCA 1080

CCAGCTGGGT [illegible] 1140

[illegible] 1200

[illegible] 1260

[illegible] 1320

[illegible] 1328

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 521 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

[illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

GAAGACCCGG CCCCGCCATA TCGATCGGCT CGCCGACTAC TTTCGTCGAA CGTGCACGCG 60

GCGGCGTCGG GCTGATCATC ACCGGTGGCT ACGCGCCCAA CCGCACCGGA TGGCTGCTGC 120

CGTTCGCCTC CGAACTCGTC ACTTCGGTGC AAGCCGACG GCACCGCCGA ATCACCAGGG 180

CGGTCCACGA TTCGTTTTCA AAGATCCTGT TCGAAATCGT CGACGCCGGA CGCTACGCCT 240

ACTACTACT TCGGTTACGA CATTTTTGCA TGAATCTGC GATCACCTTT TTTCGTCCGT 300

GATGTATT CGCTCGAAG TTTGAAGTGA CGATCTTTGA TTTCGTCTCG TGGCGCAGT 360

TGGTTCGTGA TGCTGGTGA TCATCGTCT AAATCATTCG CATCGAATCG TATCTGCTCA 420

ATGATTCAT CGTCGCGTT ACTAATAAGT TTATTGATT TCGCATCTT ACACCTGCCA 480

ACGTCGCTCT GT 492

(2) INFORMATION FOR SEQ ID NO:156:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 93 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

Gly Asp Ala Ile Val Phe Gln His Gly Asn Pro Thr Ser Ser Tyr Leu
115 120 125

Trp Arg Asn Ile Met Pro His Leu Glu Gly Leu Gly Arg Leu Val Ala
130 135 140

Cys Asp Leu Ile Gly Met Gly Ala Ser Asp Lys Leu Ser Pro Ser Gly
145 150 155 160

Pro Asp Arg Tyr Ser Tyr Gly Gln Gln Arg Asp Phe Leu Phe Asn Leu
165 170 175

Pro Asp Ala Leu Asn Leu Gly Asp His Val Val Leu Val Leu His Asp
180 185 190

Trp Gly Ser Ala Leu Gly Phe Asp Tyr Ala Asn Gln His Arg Asp Arg
195 200 205

Val Gln Gly Ile Ala Ile Met Glu Ala Ile Val Thr Pro Met Thr Trp
210 215 220

Ala Asp Trp Pro Pro Ala Val Arg Gly Val Phe Gln Gly Phe Arg Ser
225 230 235 240

Pro Gln Gly Glu Pro Met Ala Leu Glu His Asn Ile Phe Val Glu Arg
245 250 255

Val Leu Pro Gly Ala Val Leu Arg Gln Leu Ser Asp Glu Val Met Asn
260 265 270

Glu Tyr Arg Glu Pro Phe Leu Asn Pro Val Asp Arg Arg Pro Leu Trp
275 280 285

Arg Phe Pro Asn Glu Leu Pro Ile Ala Gly Glu Pro Ala Asn Ile Val
290 295 300

Ile Arg Leu Pro Ala Pro Gly Arg Asp Leu Gln Gly Tyr Gly His Gln
405 410 415

Ser Gln Pro Leu Pro Ser Gln Arg Gly Arg Gln Ile Tyr Val Ala Gly
420 425 430

Gln Arg Ser Ser Tyr Leu Pro Ser Glu Leu Val Ala Ala Phe Leu Trp
435 440 445

Ala Gln Phe Glu Gln Ala Gln Arg Ile Thr Arg Ile Arg Leu Asp Leu
450 455 460

Trp Asn Arg Tyr His Gln Pro Pro Gln Ser Leu Gln Arg Gly Leu
465 470 475 480

Leu Arg Arg Pro Ile Ile Pro Gly Gly Ser His Asn Ala His Met
485 490 495

Tyr Tyr Val Leu Leu Ala Pro Ser Arg Asp Arg Gln Gln Val Leu Ala
500 505 510

Arg Leu Thr Ser Glu Gly Ile Glu Ala Val Pro His Tyr Val Pro Leu
515 520 525

Ala Asp Ser Pro Ala Gly Arg Arg
530 535

(2) INFORMATION FOR SEQ ID NO:179:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 541 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

65 70 75 80

Gly Gly Leu Thr Val Asp Trp Lys Val Ser Trp Pro Arg Gln Arg Gly
85 90 95

Ala Thr Val Leu Ala Ala Val His Glu Trp Pro Pro Ile Val Val His
100 105 110

Phe Leu Val Ala Glu Leu Ser Gln Asp Arg Pro Gly Gln His Pro Phe
115 120 125

Asp Leu Asp Val Val Leu Gln Arg His Trp Leu Asn Leu Arg Arg Ser
130 135 140

Glu Thr Leu Glu His Thr Pro Pro Gly Arg Arg Pro Val Arg Pro Arg
145 150 155 160

His Arg Gly Asp Asp [illegible] His Arg Asn Pro Leu His Ser Val
165 170 175

Ala Met Leu Val Ser [illegible] Ala Leu Arg Arg Ala Pro Val Val
180 185 190

Gln His Gln Tyr His Val Val Ala Gln Val Gln Arg Ile Pro Glu Arg
195 200 205

Glu Gln Leu Val Ser Leu [illegible] Ile Ala Leu Ala Val Gly Pro Arg
210 215 220

Trp Ala Val Leu Val Arg Arg Ala His Pro Asp Gln Ile Ala Gly His
225 230 235 240

[illegible]

ATGAACATAT CGTCGGTCGT CGGTTCAAG ATTTTCTAC GATTCGACGT TACTCTTCC 60

GCCATGCACG CGATCGCCGG TTTCTCCGAT GCTTTGCGC AAGAGCTGCT GGTAGCGGA 120

ATCGCCGTCT CGGTGATCCA CCCGGCGCTG ACCCAGACAC CGCTGTTGGC CAACGTCGAC 180

CCCGCCGACA TGCCGCCGCC GTTTCGCAGC CTCACGCCCA TTCCCGTTCA CTGGGTCGCG 240

GCAGCGGTGC TTGACGGTGT GGCG 264

(2) INFORMATION FOR SEQ ID NO:1594:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1594:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible]

GATTACACCA ACATCGGCGC CTCCAGCCTC TCCGTGAACT ACGACCGCTA CTCCAAAAGC 1020

CGCACCACCG TCGGTTCGCA CGTACGGACC GGGTCCGACA CCATGTTCGT GGCCCCAGTA 1080

ACCATCGGCG ACGGCGCGTA TACCGGGGCC GGCACAGTGG TCCGGGAGGA TGTCCCGCCG 1140

GGGGCGCTGG CAGTGTCGGC GGGTCCGCAA C 1171

(2) INFORMATION FOR SEQ ID NO:163:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 127 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

TAAAGCTTCC ACCGCGGCCG GCCGCATCGA ATGCATTAAG TCCTTGGTTA GTCGCCCAAG 60

ACTCTGGCGA AAGCCTCAAG GGCACAACCG GCAGAATCAA GTAGGAGTGG AAGCCACGCT 120

TCTAGTC 127

(2) INFORMATION FOR SEQ ID NO:164:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

GGTC 404

(2) INFORMATION FOR SEQ ID NO:162:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1439 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

GTGGGACGCT GCCGAGCTTG TATAATAAGC AAAATATTGA TAGTTGTTCT GTTTGTGAGC 60

GGACTGAGCC ATTTAATACA GCCACTTCTA CTTTTAGGTC CGATACTCAT CTTTGGATC 120

TTATCGTGGA GTCTCTAGCAA TACTTCCTCC GAATTCTTCT TTTTTCTTAA GTGAATCGGC 180

GCTCTGAGTT CTTTACCTTC TTACTCTCA TAATGTGAT CCTCGAGTCA TGGAGCGTCT 240

CGATTGCGCC CTTCTACGAC TCGTGTGCCA TTTCTTAGCA CATCTTTTTC TAGAACGAGA 300

AGTTCATTTA TAAAAAGAC GCGACTTCCA ACGACTTCAT CATTATACCT CATCAAAGCA 360

TTAAGAAGAA TTCTCCCAAT GCGAAAATCA ATCTTCTCAT CGACGTCATT GACAACAAGC 420

[illegible] 480

[illegible] 540

[illegible] 600

[illegible] 660

[illegible] 1260

GCGGCCAGGG CGGTGCCGGT GGCAGCGGTG GCAACGGCGG CCACGGCGGC GGTGCCACCG 1320

GCGGCGCCAG CGGCAAGGGC GGCAACGGCA CCAGCGGTGC CGCCAGCGGC TCAGGCGTCA 1380

TCAACGTCAC CGCCGGCCAC GGCGGCAACG GCGGCAATGG CGGCAACGGC GGCAACGGC 1439

(2) INFORMATION FOR SEQ ID NO:168:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible] acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible]

[illegible]

(2) INFORMATION FOR SEQ ID NO:169:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:
(B) TYPE:
(C) STRANDEDNESS:
(D) TOPOLOGY:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 392 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible] 300

[illegible] 360

[illegible] 392

(2) INFORMATION FOR SEQ ID NO:166:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

[illegible] 540

(2) INFORMATION FOR SEQ ID NO:167:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 690 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

[illegible]

[illegible] 60

[illegible] 120

[illegible] 180

GACGAGCTAC GCCGATGTGG GAACGTCGTT [illegible] 240

[illegible] 300

[illegible] 360

[illegible] 420

[illegible] 480

[illegible] 540

[illegible] 600

[illegible] 660

[illegible] 60

TGGGTACAGT CGGGGTTGGC GGCTACGGTG GCTCCTCTAC TACGTTATTT GCACTTGGCG 120

GCTCTGGGGG CACCGGCGGC GACTGGCGGA CCGGCGGGCG TGGCGGCTTG TTAATGGGCG 180

CCGGCGCCGG CGGGCACGGT GGCACTGGCG GCGCGGGCGG TGCCGGTGTC GACGGTGGCG 240

GCGCTGCGGT GGCCGGCGGG CCCGGCGGCA ACTGCGGCGC CGGGGGTCAA GCCGCCCTGC 300

TGTTCGGTCT CGCCGTTACC GATGAGTCTT GTTTCTACGA TGTGGATGTT GGTGCGGGCG 360

GTTATTCTTT CTATTTACTC ATTCTTGCCT TTTTTGTTAC TTTTGGC 407

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] pairs
(B) TYPE: [illegible]
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible] 300

(x.) SEQUENCE DESCRIPTION: SEQ ID NO:170:

GGTGGTAACG GCGGCCAGGG TGGCATCGGC GGCGCCGGCG AGAGAGGCGC CGACGGCGCC 60

GGCCCCAATG CTAACGGCGC AAACGGCGAG AACGGCGGTA GCGGTGGTAA CGGTGGCGAC 120

GGCGGCGCCG GCGGCAATGG CGGCGCTGGC GGCAACGCCG AGGCGGCCGG GTACACCGAC 180

GGCGCCACGG GCACCGGCGG CGACGGCGGC AACGGCGGT 219

(2) INFORMATION FOR SEQ ID NO:171:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 494 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible] 300

[illegible] 360

[illegible] 420

[illegible] 480

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GGGCCGGTGG TGCCAAGCCC CATCTCTTCA GCCTGGACGG CGCGGCGGCT CCCTTGGGGC 60

TTGGCGGCAC CGCCGGCCAG GGTGGGGCTG GCCGTGCCGG AGCGGCCGGC GCCGACGCCC 120

CCGCCAGCAC AGGTCTAACC GGTGGTACCG GGTTCGCTGG CGGGGCCGGC GGCGTCGGCG 180

GCCAGAGCGG CAACGCCATT GCCGGCGGCA TCAACGGCTC 220

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 398 bp
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ACCCCCCAA CCCCCCCCCC GGCCCCCCCC CCCCCCCCCC CAACCACAAC CCCAACGGGC 60

CCCCATCCCC CCCCCACCCC CCAACCCCCC CACCCCCCCC CCCCCGCCCC AAGCACCACC 120

CCACCCCCCC ACCCCCACCC CCAACCCCCC ACCCCCACCC CCACCCCACC CCCCCCCTCC 180

CCCCCTCGCC CACCCCCCCC ACCACCCCCA CCCCCCCCCC CCCCCCCCCA ACCCCCCCCA 240

ACCCCACCCC CCCCCCCCCC CCACCCCCCC CAACCCCCCC ACCCCCCCCC CCCCCACCCC 300

CCCCCACCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCAC 350

CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC 380

[illegible] 19

GCGAAAACGG CGGAAACGGC GCAGACAACA CCACCACCGC CGCCGCCGGC ACCACAGGCG 240

GCGACGGCGG GGCCGGCGGG GCCGGCGGAA CCGGCGGAAC CGGCGGAGCC GCCGGCACCG 300

GCACCGGCGG CCAACAAGGC AACGGCGGCA ACGGCGGCAC CGGCGGCAAA GGCGGCACCG 360

GCGGCGACGG [illegible] 400

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible]
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: [illegible]

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:[illegible]:

[illegible] 60

[illegible] 120

[illegible] 180

[illegible] 240

[illegible] 300

[illegible] 360

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

GAGCCGGTGG TGCCCCGGGC CAGTTCTTCA GCGCCGAGGG CGCGCAGGGT GCCTTTGCGG 60

TTGGCGGCAC CGGCGGCCAG GGTGGGGCTG GCGGTGCCGG AGCGGCCGGC GCCGACGCCC 120

CCGCCAGCAC AGGTCTAACC GGTGGTACCG GGTTCGCTGG CGGGGCCGGC GGCGTCGGCG 180

GCCACGGCGG CAACGCCATT GCCGGCGGCA TCAACGGCTC CGGTGGTGTC GACGGCACC 239

(2) INFORMATION FOR SEQ ID NO:177:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

[illegible]

CGTACTGCCC CAAACACTCA AAAA 44

(2) INFORMATION FOR SEQ ID NO:178:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2138 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

CTTTCCTATT ATTCGTATTT TTTTTATTT TAATTTTTAA ATTCATTTTT TTTCAAATAC 60

GTAGCAAAAT TTTTATCAAA TGGTATTTT GAAGCATTTT GATAGTTTTT TTAGCAGTTG 120

ATAAACGGGT GGCATCAAAT TATTCATTA ATCTTTTATT TTATGTATTA TAATTGTTAT 180

AGCACTAACG AGCATGATCT GATATGACTT ATTTGAAGAC TCTACTGTG GATCAGCAAG 240

AGATTTTTAA CAGCTTAAAT CAGTTCATAT TTTTATGTTC TATCGTTCTT ATTGATGTCC 300

TTATCACATT GTCGTAACTT ACGGTGGTTA AAAACTGTTT TAACTAGTTT TTATTTTCCG 360

TCTAAAAAAT GTATTAAATA TTTGTTAAAA TCTTTAAAAA ATTTTATTTT TTTTTGACTT 420

TCTTTTTAA TTTTTTTAT TTTATTTTTA ATTTTTTTTA TTTTTTATTT TTTTTTTTTA 480

ATAATTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 540

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 600

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 660

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 720

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 780

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 840

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 900

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 960

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1020

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1080

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1140

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1200

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1260

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1320

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1380

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1440

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1500

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1560

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1620

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1680

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1740

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1800

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1860

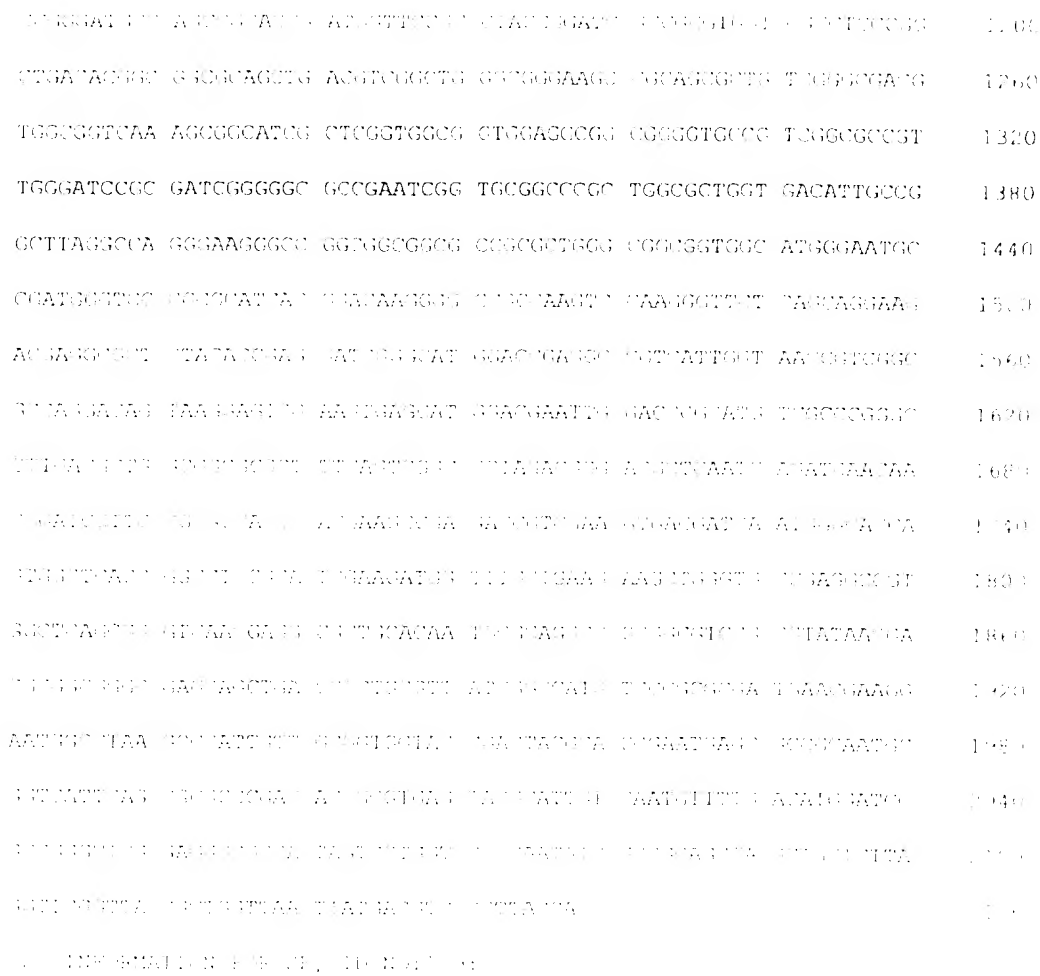
TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1920

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 1980

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 2040

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 2100

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 2138



35 40 45

Leu Val Leu Ser Ala Asp Asn Met Arg Gln Tyr Leu Ala Ala Gly Ala
50 55 60

Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Ala
65 70 75 80

Tyr Gly Glu Val Asp Glu Glu Ala Ala Thr Ala Leu Asp Asn Asp Gly
85 90 95

Glu Gly Thr Val Gln Ala Gln Ser Ala Gly Ala Val Gly Gly Asp Ser
100 105 110

Ser Ala Glu Leu Thr Asp Thr Pro Arg Val Ala Thr Ala Gly Glu Pro
115 120 125

Asn Phe Met Asp Leu Lys [illegible] Ala Ala Arg Lys Leu Glu Thr Gly Asn
130 135 140

Gln Gly Ala Ser Leu Ala His Phe Ala Asp Gly Trp Asn Thr Phe Asn
145 150 155 160

Leu Thr Leu Gln Gly Asp Val Lys Arg Phe Arg Gly Phe Asp Asn Trp
165 170 175

[illegible] Gly Asp Ala Ala Thr Ala Tyr [illegible] Ala [illegible] Leu Asn Gln Glu Arg
180 185 190

Gln Thr [illegible]
195 200 205

[illegible]

Ala Gln Leu Thr Ser Ala Gly Arg Gln Ala Ala Ala Leu Ser Gly Asp
340 345 350

Val Ala Val Lys Ala Ala Ser Leu Gly Gly Gly Gly Gly Gly Gly Val
355 360 365

Pro Ser Ala Pro Leu Gly Ser Ala Ile Gly Gly Ala Glu Ser Val Arg
370 375 380

[illegible]
385 390 395 400

[illegible]
405 410 415

[illegible]
420 425 430

[illegible]
435 440 445

[illegible]
450 455 460

[illegible]

[illegible]

Arg Asp Glu Ser Leu Leu Leu Arg Arg Arg Gly Arg Val Asp Leu Asp
100 105 110

Gly Gly Gly Arg Leu Arg Arg Val Tyr Arg Phe Gln Gly Cys Leu Val
115 120 125

Val Val Phe Gly Gln His Leu Leu Arg Pro Leu Leu Ile Leu Arg Val
130 140

His Arg Glu Arg Leu Val Ala [illegible] Arg Arg Val Phe Arg Val Lys Pro
145 155 160

Phe Glu Pro Asp Tyr Val Phe Leu Ser Arg Met Phe Pro Pro Ser Pro
165 170 175

Met Val Leu Leu Arg Asp Leu Leu Ser Leu Leu Gly His Arg Ser Ala
180 185 190

Gln Phe Gly His Val Glu Tyr Pro Leu Pro Leu Leu Ile Gln Arg Ser
195 200

Leu Ala Ser Gly Ser Arg Leu Ala Pro Pro Val Val Lys Pro Pro Gln
210 215 220

Pro Leu Asp Val Ala Leu Gln Arg Leu Val Glu Ser Val Pro Pro
225 230 235 240

Arg Cys Val Arg [illegible] Arg [illegible] Gln Arg Arg Leu Pro
245 250 255

[illegible]

Leu Asp Pro Ser Ser Val Ser Ser Leu Leu Asn Asn Val Ser Cys Gly
20 25 30

Ala Glu Ser Ser Ala Ser Ser Ser Ala Arg Ser Gly Asn Gly Ser Arg
35 40 45

Trp Thr Ser Met Pro Ser Gly Thr Arg Pro Gly Pro Arg Arg Ala Thr
50 55 60

Ser Arg Asp Asp Arg Arg Ser Ala Thr Ser Val Ile Pro Ser Arg Arg
65 70 75 80

Ser Val Ala Pro Arg Ala Gly Pro Gly Thr Arg Leu Ala Ser His Arg
85 90 95

Ala Ser Leu Ser Asn Ala Cys Leu Val Arg Leu Val Ser Ser Ala Ser
100 105 110

Gly Arg Ile Leu Ser Ser Leu Ile Leu Val Arg Ser Arg Cys Val
115 120 125

Asp Lys Asn Gly Arg Arg Cys Ala Ser Gly Cys Arg Arg Leu Asn Arg
130 135 140

Ala Arg Ser Ser Ser Ile Ala Arg Ser Cys Arg Thr Ile Gly Thr Phe
145 150 155 160

Arg Arg Ser Arg Cys Tyr Ala Asn Ser Ser Arg Val Ser Thr Asn Ser Ile
165 170 175

His Val Ser His Gly Val Ala Leu Gly Thr Ser Arg Leu Ile Gly Gly
180 185 190

35 40 45

Arg Asn Pro Arg Arg Ser Ser Arg Arg Asp Ala Glu Asp Arg Arg Val
50 55 60

Ile Phe Ala Ala Thr Leu Val Ala Val Asp Pro Pro Leu Arg Gly Ala
65 70 75 80

Gly Gly Glu Ala Asn Gln Leu Ile Asp Leu Gly Val Cys Arg Arg Glu
85 90 95

Ala Gly Arg Val Arg Arg Tyr Tyr Leu Leu His His Arg His Arg His
100 105 110

Gln Gly Ala Ala [illegible] Asp [illegible] Arg Arg Arg Arg Arg His Arg Arg Val
115 120 125

Gln Val His Arg Arg [illegible] Val Arg Gln Leu Arg Arg Tyr Val
130 135 140

Gln Thr Ala His His Arg Arg Pro Ala Arg Thr Asp Arg Val Arg His
145 150 155 160

His Val Arg Tyr Pro Ser Asp Ser Arg Arg Arg Val Tyr Arg Gly
165 170 175

Arg His Ser Gly Arg Gly [illegible] Gly Gly Ala Gly Ser Val
180 185 190

Gly Gly Ser Ala
195

[illegible]

Ala Arg Gly Gly Tyr Asp His Pro Ile Leu Arg Val Thr Pro Gly
50 55 60

Ala Thr Pro Trp Val Thr Trp Gly Glu Phe Val Glu Thr Arg Met Leu
65 70 75 80

Ala Glu Tyr Arg Asp Arg Arg Lys Val Pro Ile Val Arg Gln Arg Ala
85 90 95

Ala Ile Glu Ala Leu Arg Ala Arg His Asn Leu Arg Tyr Pro Leu Ala
100 105 110

His Leu Ala Ile Glu Ile Leu Ser Thr His Glu Arg Asp Leu Thr Met Gly
115 120 125

Pro Glu Gly Ile Leu Ala Leu Ile Asp Ala Tyr Val Thr Ile Arg Thr Gly
130 135 140

Ile Ala Leu Ile Ile Gly Arg Arg Thr Leu Ala Leu Leu Val Pro Asn
145 150 155 160

Ser Ala Arg Gly Ala Thr Leu Arg Arg Leu Gly Ile Thr Asp Val Ala
165 170 175

Asp Leu Arg Ala Ile Pro Arg Glu Thr Ala Arg Arg Leu Gly Arg Val
180 185 190

Thr Asp Gly Leu Asp Val His Leu Ile Phe His Arg Leu Asn Asp
195 200 205

Asp Arg Asn Asp Ala Leu Ala Leu Thr Ala Ile Tyr Arg Leu
210 215 220

Ile His Leu Asn Gly Leu Leu Gly Arg
225 230

Asp Ala Met Arg Leu Asp Met Ile Leu Ser Ala
235 240

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

TTGGCTTCTT CTTACTTCTT CAAGCTAACT TTCAAAGCC TTGCACTAA TAACGATGG 1440

AATCGGGGGT GCATATCATC CCACAGGGCC TTGGTGCCAT GCTGGCGATG CCGATCGCCG 1500

GAGCGATGAT GGACCGACGG GGACCGGCCA AGATCGTGCT GGTTGGGATC ATGCTGATCG 1560

CTGCGGGGTT GGGCACCTTC GCCTTTGGTG TCGCCGCGCA AGCGGACTAC TTACCCATTC 1620

TGCCGACCGG CTTCGCAATC ATTTCATGG CATCGGTTG CTCCATCATC CCACTGTCCG 1680

GCTCGGCAGT TAGCCCCTG CTACCATC AGATCCTTCC CGTTTCGATC TGATCAGCG 1740

TCAACCTACA CTTCGTTCT TCGATAGCTA CTTCGTCGAT GTCTTCTTC TTCACCTACC 1800

ACTTTAATCA TA..... ATTTTACTT CAAAAAACTT TCCATTAAT TAGAGAGTTC 1860

GTCCTCGCG CGCCTCTTC GTCATCCCC ATTCTTAAC TCCCTAAAA AACTTTGCGC 1920

CCCAACTGCT CTATGACTTT TCGCACGCCT ATGAGGTTCT ATTCGTAATA CTGGTCGCGC 1980

TACTGGTCCT GACCTTAAT CGCTCGTTAT TCTTCCTCAA CAAGTAGCTT ACTCATCTAA 2040

CTACACTCTT CTATCTCCA TGATCTTCTC TT 2072

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
A) LENGTH: 102 base pairs
B) TYPE: nucleic acid
C) STRANDEDNESS: single
D) TOPOLOGY: linear

GAACATCGCC ACGGCAATCA CGACCACTAG ATTTCCCACG CATACCCTTT CGTACCGGTG 600

CGCCGCGGTT GGTCGATCGG TCGCATATCG ATGGCGCCGT TTAACGTAAC AGCTTTCGCG 660

GGACCGGGGG TCACAACGGG CGAGTTGTCC GGCCGGGAAC CCGGCAGGTC TCGGCCGCGG 720

TCACCCCAGC TCACTGGTGC ACCATCCGGG TGTCGGTGAG CGTGCAACTC AAACACACTC 780

AAGGGCAACG GTTTTTCAGG TTACCAGTTT AAGTTGGACC CGCAATCGCT CGTACGTTTC 840

GATCGGCGCT AGGTTCTCGA TCGCCATCTT TTTCATCGTA GTTTTCGCGC TGAAGCCGAC 900

GATGCATCGC TATTTCTCCT CATCGTCGAC ATCGTCGTCG GCGAGGTCGT CGGTCAAGCC 960

GCGATCGGAG TAGCGATCGA TCTCGTAGCTA TTCGTTCGAC ATCATGTCTC CGAAGTAGGC 1020

GTCGAGCTCA AAGAGCCAAA CACGATCGAA TTCGTCGTAT TCGTCGAAAT AGTC 1140

TTTCACAACG CGTTTCAGGA TCAAATCGAC TCGAATCTTT ACTCGCGTCT CGCGGATCAC 1200

CGCTTCGTCT GATCGAGCGT TCCCGAACGC ATTCTCGTCG TCAACGGCGG CATCGCACGT 1260

TTCGCCCCCG CTAAAAAAT CGATCTTTT CGAATAATT TACACTTCGT TTGGGCAGGT 1320

TTAACTCGA TCGCTCGAT TATCTCGCG ATTCAATTT TCTCGTTCGT CTCGAAGCCA 1380

AGTTCTTCAA GAACTTCTTC GCAATTATA AACTTTTCGA TATCGACTTT CTTCTTCTGC 1440

ATTTCAAAA ATTTATTAT TGATGTTTTA TATTTTCTCA TTTTTCAAA 1500

ATTTTTTA TCGTTCGCA ATCTTTTTT CCTCTTCGTT TCGTTTTTT 1560

AAAAATCGT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT 1620

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

CTGGCTGCT AGTGTCACCG GCGATATGAC ATCGCATTC AATTTCGCGG CCCGCCGGGA 60

CCGTTGCCA TCAATCTG ACGACCCTT CC TCAATTC CCGAAGTCC CAAGTGCGT 120

CCCAATGTG GTGTGGTTT TCTTAAACA ATCTTCCG TATTGGCC TTACCCCCA 180

AAAAACCCA GTCACAAAT CCCTCCCC CCCCGATT CACGCCA TCTGCTAGCC 240

GGCATGTTT TAGAGCTAA GCTTGCTAG TCAAAA CAC AGCC CATCGACGG 300

GTCATGCCA GCGAAATTAG AAAATCCGG ATTTTCTCC GCATTGCCC TAGATGCTG 360

ATCCTTGGT CGTTCGTCTT TAACGATTA CTCTCAGCC TTAAACCC TGCCTGGAA 420

GGCGACAACG TGTTTTGCC TCTTCGTCCG CTTTTCCC CAACACCCC GCTGACGAAA 480

CCCCTGAGT CGCGAATGCT TGCTTTCT CTCCCCCC CCGCTGCGC CCCCTCTCT 540

CTACCCATG ATTATCGCCA CTCGGCCCC CCTCAACGC ACTCCCCCA TCTCGATGTC 600

CATCCCCCC AACGCCCCA CCCCACCC CCCCCCCC CCCCCCCC CCCTCGCAC 660

CCCCCAAC ACTCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC 720

ACCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC 780

CCCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC 840

CCCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC CCCCCCCC 900

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

CCGGCTGGTT GTTGGCATAC TCCGCCGCGG CGGCCTCGAC CGCACTGGCC GTGGCGTGTG 60

TCCGGGGCTGA CCACCGGGAT CGCCGAACCA TCCGAGATCA CCTCGCAATG ATCCACCTCG 120

CGCAGCTGGT CACCCAGCCA CGGGGCGGTG TGCGACAGCG CCTGCATCAC CTTGGTATAG 180

GCCTCGGCGC CGACGGGCAG GAACTTGTAG TACTCGCCGA CCACCTGGTT ACCGGGACGG 240

GAGAAGTTCA GGCTGAAGCT CGGCATGTAT CCTCGAGCGT AGTTGACCTG GAAAACTAGA 300

TCTTTTGGCA GGTGCTCGGG CGCTGCCAC AGGACAAACC CGACGCCGGG ATAGGTCAG 359

(2) INFORMATION FOR SEQ ID NO:188:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 920 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

Gln Gln Pro Lys Gly Pro Phe Gly Gln Val Ile Gln Ala Phe Ala Asp
1 5 10 15

Gly Leu Ala Gly Lys Gly Lys Gln Ile Asn Thr Thr Leu Asn Ser Leu
20 25 30

Ser Gln Ala Leu Asn Ala Leu Asn Gln Gly Arg Gly Asp Phe Phe Ala
35 40 45

Val Val Asn Ser Leu Ala Leu Phe Val Asn Ala Leu His Gln Asp Asp
50 55 60

Thr Gln Phe Val Ala Leu Asn Lys Asn Leu Ala Gln Pro Thr Asp Arg
65 70 75 80

Leu Thr His Ser Arg Ala Arg Leu Ser Asn Ala Ile Leu Phe Asp
85 90 95

Ser Leu Leu Ala Val Ala Arg Ile Phe Pro Ala Tyr Arg Arg His Val
100 105 110

Leu Thr His Asp Val Asn Asn Leu Ala Arg Val Thr Thr Leu Leu
115 120 125

Thr Pro Asp Leu Asp Gly Leu Val Leu His Ile Phe Pro
130 135 140

Thr Leu Ala Ala Asn Asn Asn Leu Tyr His Leu Thr His Gly Gly
145 150 155 160

Val Val Leu Ala Phe Thr Asn Ile Ala Asn Leu Ser Ala His
165 170 175

Ala Val Leu Ala Arg Ala Val Leu Arg Phe Val Leu
180 185 190

Ser Leu Ala Gln Leu Met Gly Gly Pro Asp Ile Ala Pro Pro Ser Ser
290 295 300

Gly Leu Gln Thr Pro Pro Gly Pro Pro Asn Ala Tyr Asp Glu Tyr Pro
305 310 315 320

Val Leu Pro Pro Ile Gly Leu Gln Ala Pro Gln Val Pro Ile Pro Pro
325 330 335

Pro Pro Pro Gly Pro Asp Val Ile Pro Gly Pro Val Pro Pro Val Leu
340 345 350

Ala Ala Ile Pro Pro Pro Arg Asp Arg Pro Ala Ala Ser Glu Asn Phe
355 360 365

Arg Tyr Met Gly Gln Leu Leu Leu Ser Pro Gly Leu Thr Phe Leu
370 375 380

Ile Gly Val Ser Ser Ser Pro Ala Arg Ser Thr Met Ala Asp Arg His
385 390 395 400

Val Leu Ile Pro Ala Leu Thr Gly Leu Ala Leu Ile Ala Ala Phe Val
405 410 415

Ala Ser Ser Trp Tyr Arg Thr Arg Ser Pro Leu Ile Asp Met Arg Leu
420 425 430

Ile Gln Arg Arg Ala Val Ala Gln Ala Met Thr Met Thr Val Leu
435 440 445

Leu Leu Ala Leu Phe Ser Ile Ser Ile Gly Ser Pro Leu Gln
450 455 460

Ser Thr Leu Ala Ser Pro Ala Ala Ala Pro Pro
465 470 475

Gly Gly Arg Pro Ala Ala Ala Ala Pro Pro
485 490

Gln Phe Asn [illegible] Ser Glu Ile Ile Ala Thr Ala Lys Lys Val Ala Leu
595 600 605

Thr Pro Glu Ser Gly Ala Gly Arg Gly Ala Ala Val Asp Pro Ser Ser
610 615 620

Leu Pro Arg Gln Thr Asn Phe Ala Ala Gln Leu Leu His Asp Leu Ser
625 630 635 640

[illegible] Ala Lys Ala Val Val Phe Val Ile Ala Thr Ala Leu Val Val Ser
645 650 655

Thr Leu Ile Pro Ala Ala Phe Leu Ile Lys Gln Glu Ala Ser His Arg
660 665 670

Ala Ala Pro Leu Leu Ser Ala
675

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO:[illegible]:

[illegible]

[illegible]

Thr Arg Arg Asp Ile Arg Thr Ala
115 120

(2) INFORMATION FOR SEQ ID NO:191:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 89 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

[illegible]

[illegible]

His Leu Ala Met Ile His Leu Ala Gln Leu Val Thr Gln Pro Pro Gly
35 40 45

Gly Val Arg Gln Arg Leu His His Leu Gly Ile Ala Val Ala Pro Gln
50 55 60

Pro Gln Glu Val Val Val Leu Ala His His Leu Val Thr Gly Thr Gly
65 70 75 80

Glu Val Gln Gly Glu Gly Arg His Val Ala Ala Glu Val Val Asp Pro
85 90 95

Thr Asn Gln Ile Leu Arg Gln Val Leu Gly Pro Ala Pro His Asp Lys
100 105 110

Ile Arg Arg Arg Gly Leu Gly Gln
115

(2) INFORMATION FOR SEQ ID NO:1964:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 114 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1964:

Arg Ala Arg Ile Pro Arg Arg Leu Gly Gln Thr Arg Leu Val Leu Ala

Thr Leu Arg Gly Gly Ala Arg Arg Leu Gly Ala Leu Val Asp

INFORMATION FOR SEQ ID NO:194:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 811 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

[illegible]

[illegible]

GACTTTGTGG TCCCGGTGGC GGGATAGAGC ACCTGTCAGC GTTGGTCAGC GTCACCCGTT 120

GCTCGGACGC CGAACCCATG CTTTCAACGT AGCCTGTAGG TCACACAAGT CGCGAGCGTA 180

ACGTCACGGT CAAATATCGC GTGGAATTTC GCCGTGACGT TCCGCTCGCG GACAATCAAG 240

GCATACTCAC TTACACCGGA GCTATTTGTA CGTTTCGAT CGTTTCCGCT CTGGTGAACG 300

TCCCGTCAAA TGTGTATACT CTTACAGTAG ACAACGACAT TACGTTCTAT CAGGTGCACG 360

CTAAAGACAA CGGACGTATC CCTTACAAGC CGCTCTGCGA GGCCTGTCGT GAGGTGGTCG 420

ATACGCCAAT TTTCGTTGTT TTTATCATAT TCTTTGATTC CAAATGTTTT CGATCACCG 480

ATTCGACGAT ATTTATTTTT TTGAAGAAT TATTTTTGA ATCTGAGTTT TTGAGTTCT 540

TGTAGCCGCT GGACTTGAGT CGGATGATCT TCTGCTCGAG CTATTTTTTG GAGCTTGATT 600

CTGACTATTC TAAATCTATT TTCTTCTGT CTAAAAATAT CTTGAGACCT ACCCGATGG 660

AGTCGTCGA TTCCTCAGCT TTTCTGAAAT CAATAAAAAA TAAGAGCTGT TATCCACAAT 720

TCGTCTCGA CTTCGCTTAT ACAAAGTTCT ATGATGATTC AGATTTCGCT GAGTTGAAGT 780

CTTTTTATAT ATGAATCTCT TTCTTGATT ATCATAGTTA AAATACTTTT ATTCATGTCA 840

TTGTTATTTT TAACAAATAA TAATTTTTT ATTTTTTTAA ATTTTTTTTA TTTTTGACTT 900

AATTTTTAAT TTTTTTAAAA TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT 960

AAAAAAAAAA 970

CTCCGCCGCT GCCACCGTTG [illegible] [illegible] [illegible] CCACCCTTGC 240

CACCCACGCC ACCGGCTCGG CCCACCCCGC CGACACCAAG CGAGCTGCCG CCGGAGCCAC 300

CATCACCACC TACGCCACCG ACCGCCCAGA CACCAGCGAC CGGGTCTTCG TGAAACGTCG 360

CGGTGCCACC ACCGCCGCCG TTACCGCCAA CCCCACCGGC AACGCCGGCG CCGCCATCCC 420

CGCCGGCCCC GGCGTTGCCG CCGTTGCCGC CGTTGCCGAA CAACAACCCG CCGGCGCCGC 480

CGTTGCCGCC [illegible] [illegible] [illegible] [illegible] CTGCCGCCCT 540

[illegible] [illegible] [illegible] [illegible] [illegible] TCTGGGCTGT 600

[illegible] [illegible] [illegible] [illegible] CCCCCCCGTG GCGCCGTCAC 660

[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]

[illegible] [illegible] [illegible] [illegible] [illegible] CCGCACTGG 780

[illegible] [illegible] [illegible] [illegible] [illegible] CCGATGTGTA 840

CGAAAGCCCC [illegible] CCGCCGCTAC [illegible] [illegible] ACACCGTCGG 900

[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]

[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] 1020

[illegible] [illegible] [illegible] [illegible] [illegible] ACACCGCTCG 1080

CGGCTGTATT AGCTGTCGTC ACTTGGAAC TAATATGCA TAAACTTAAA CATCAGTTTC 1860

CTAGCAGCTG CGCGTCGCGC ATCATTAAGC ATACTTTGCA CTTCTGGATA TCTATATCG 1920

CCGCACCGTG TCCCCAGCGG CCACGTGACC TTTGGTCGCT GGCTGGCGGC CCTGACTATG 1980

GCCGCGACGG CCCTCGTTCT GATTCGCCCC GGCGCGCAGC TTGTTGCGCG AGTTGAAGAC 2040

GGGAGGACAG GCCGAGCTTG CGCTAGACGT GGGTCAAGTG CGAATGCAGA CTCCCGGCGG 2100

AGATGAATAG GCGGACGCGG ATCTCGTTGT TCGTGAGTCC CTCACTGACT AGTAGAGCCA 2160

CTTCAACGTC TTTCGCCTTT AACGCGCCCT AGCGCTTTCT CCGCCTTTC TTCGACCGC 2220

GGCCTTTTG CGCTTAGCGC ATCGCCTTAT CTAICGATAA CGAAGTTCAT TCTGCCCAGG 2280

GATCTCTAAA CTCTTTTTGA CATATGTACT CTTCGAAGGT TCTTACTTTA TATTACAGCT 2341

TTTTTTTAA AGCTGAAA AAAA 2365

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 376 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

Gly Gly Thr Val Ala Ala Gly Ala Thr Gly Arg Arg Ala Gly Ser Ala
115 120 125

Met Ala Ala Arg Ala Ala Val Ala Ala Gly Leu Ile Thr Asp Ala Gly
130 135 140

His Ile Cys Arg Ala Val Pro Gly Ala Gly Arg Gly Ala Gly Arg Gly
145 150 155 160

Ile Asp Pro Val Cys Pro Gly Gln Ala Gly Ala Ala Gly Thr Thr Gly
165 170 175

Ala Ala Met Ala Gln Gln Pro Gly Val Ala Ala Val Thr Ala Arg Thr
180 185 190

Pro Asp Ala Tyr Gly His Ala Gly Ala Ala Asp Thr Ala Val Ala Ala
195 200 205

Val Ala Pro Gln Pro Pro Pro Val Pro Thr Gly Thr Ala Gly Arg Ala
210 215 220

Gly Thr Thr Gly Pro Ala Val Ala Ala Val Ala Asp Gln Pro Gly Arg
225 230 235 240

Ala Ser Ala Ala Ala Gly Leu Thr [illegible] [illegible] Ala Ser Arg Ala Val Ala
245 250 255

Thr Val Ala Cys Thr Gln Pro Ala Gly Arg Ala Arg Leu Pro Gly Cys
260 265 270

Arg Pro Val [illegible] Ala Val [illegible] Arg [illegible] [illegible] Ala Pro Gln [illegible] Arg Leu
275 280 285

[illegible] Pro

[illegible]

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

```
TTCAATATCG GCATCGGAAA CATCGGAGTA TTCAATGTCG GTTTCGGAAG CGTCGGAAAC   1440
TACAACATCG GATCCGGAAA CGTCGGGATC TACAACATCG GTTTTGGAAA CGTCGGCGAC   1500
TACAACGTCG GCTTCGGGAA CGCGGGCGAC TTCAACCAAG GCTTTGCCAA CACCGGCAAC   1560
AACAACATCG GGTTCGCCAA CACCGGCAAT AACAACATCG GCATTCGGCT GTCCGGCGAC   1620
AACCAGCAGG GCTTCAATAT TGCTAGCGGC TGGAACTCGG GCACCGGCAA CAGCGGCCTG   1680
TTCAATTCGG GCACCAATAA CGTTGGCATT TTCAACGCGG GCAACGGAAA CGTCGGCATC   1740
GGAAATTCGG GCACCGGGAA [illegible] [illegible] GCACCGAGAA TACCGGCATC   1800
TTCAATGCTG GCACCTACAA CACCGGCATT TTCAACGCCG GCGACTTCAA CACGGGCTTC   1860
TACAATACGG GCAGCTACAA CACCGGCGGT TTCAACGTCG GCAACACCAA CACCGGCAAC   1920
TTCAATGTTG GTGACACCAA TACCGGCAGT AATAACGTCG GGGACACCAA CACCGGCTTT   1980
TTCAATGCTG GCAATGTCAA TACCGGCTTT TTCAATACCG GCGACTTCAA CAATGGCTTC   2040
TTGGTGCCGG GCGATAACCA GGGCCAGATT GCTATCGATC TCTCGGTCAC CACTCCATTC   2100
ATTCCTATAA ACGAGCAGAT [illegible] [illegible] TAATGA[illegible] [illegible]   2160
[illegible] TCACCGGACC [illegible] [illegible] [illegible] GAGCGGTTTG   2220
[illegible]   2280
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
```

(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Gly Gln Asn Ala Pro Ala Ile Ala Ala Thr Glu Ala Ala Tyr Asp Gln
1 5 10 15

Met Trp Ala Glu Asp Cys Ala Ala Met Ile Gly Tyr Ser Ala Gly Ala
20 25 30

Ser Ala Ala Val Ser Ala Leu Thr Ile Gly Glu Ala Leu Thr Thr
35 40 45

Val Asn Gly Gly Gly Ala Leu Cys Ser Ala Ala Ala Gln Val Thr
50 55 60

Thr Ala Val Pro Arg Asn Leu Gly Leu Ala Asn Val Arg Asn Gly Asn
65 70 75 80

Val Ala Ala Gly Asn Val Ala Ala Ile Phe Asn Leu Gly Asn Ile
85 90 95

Gly Asn Gly Asn Thr Gly Asn Ile Gly Asn Ile Gly Asn Ile Gly
100 105 110

Phe Gly Ala Val Gly Ile Gly Leu Ile Ala Ala Asn Ala Gly
115 120 125

Ile Ala Thr Gly Gly Asn Ala Ile Gly Ile Gly Ala Ala
130 135 140

Gly Ser Tyr Asn Pro Gly Asn Ser Asn Thr Gly Gly Phe Asn Met Gly
245 250 255

Gln Tyr Asn Thr Gly Tyr Leu Asn Ser Gly Asn Tyr Asn Thr Gly Leu
260 265 270

Ala Asn Ser Gly Asn Val Asn Thr Gly Ala Phe Ile Thr Gly Asn Phe
275 280 285

Asn Asn Gly Phe Leu Trp Arg Gly Asp His Gln Gly Leu Ile Phe Gly
290 295 300

Ser Pro Gly Phe Phe Asn Ser Thr Ser Ala Pro Ser Ser Gly Phe Phe
305 310 315 320

Leu Ser Gly Ala Gly Ser Asn Ser Gly Phe Leu Asn Ser Gly Ala Asn
325 330 335

Asn Ser Gly Phe Phe Asn Thr Ser Ser Gly Ala Ile Gly Asn Ser Gly
340 345 350

Leu Ala Asn Ala Gly Val Leu Val Ser Gly Val Ile Asn Ser Gly Asn
355 360 365

Thr Val Ser Gly Leu Phe Asn Ser Ser Leu Val Ala Ile Thr Thr Pro
370 375 380

Asn Ser Ile Ser Gly Phe Phe Asn Thr Gly Ser Asn Met Ser Gly Phe
385 390 395 400

Phe Gly Ser Ser Ser Phe Asn Ser Phe Asn Leu Gly Ile Ala Asn Arg Gly Val
405 410 415

Ser Ser Leu Phe Ala Ile Gly Ser Gly Asn Leu
420 425

Ser Gly Asn Val Gly Arg Ile Ala Leu Gly Ser Gly Asn Leu
435 440 445

530 535 540

Phe Asn Ile Ala Ser Gly Trp Asn Ser Gly Thr Gly Asn Ser Gly Leu
545 550 555 560

Phe Asn Ser Gly Thr Asn Asn Val Gly Ile Phe Asn Ala Gly Thr Gly
565 570 575

Asn Val Gly Ile Ala Asn Ser Gly Thr Gly Asn Trp Gly Ile Gly Asn
580 585 590

Pro Gly Thr Asn Asn Thr Gly Ile Leu Asn Ala Gly Ser Tyr Asn Thr
595 600 605

Gly Ile Leu Asn Ala Gly Asp Phe Asn Thr Gly Phe Tyr Asn Thr Gly
610 615 620

Ser Tyr Asn Thr Gly Ile Asn Val Gly Asn Thr Asn Thr Gly Asn
625 630 635 640

Ile Asn Val Gly Asp Trp Asn Ile Gly Asn Ile Asn Asn Thr Gly Asp
645 650 655

Asn Thr Gly Ile Phe Asn Ile Gly Asn Val Asn Thr Gly Ala Phe Asp
660 665 670

Thr Gly Asp Phe Asn Ala Gly Phe Leu Ala Gly Asn Asn Gln Gly
675 680 685

Val Ile Ala Asp Ile Asn Val Thr Gly Ile Asn
690 695

Ile Ser Phe Ala Gly Asn
700

Ile Ile Ala Thr Gly Asp Ile Asn Phe Ile Gly Ile
705 710

Ala Ile Gly Asn Ser Gly Phe Gln Asn Leu Gly Ser Leu Gln Ser Gly
835 840 845

Trp Ala Asn Leu Gly Asn Ser Val Ser Gly Phe Phe Asn Thr Ser Thr
850 855 860

Val Asn Leu Ser Thr Pro Ala Asn Val Ser Gly Leu Asn Asn Ile Gly
865 870 875 880

Thr Asn Leu Ser Gly Val Phe Arg Gly Pro Thr Gly Thr Ile Phe Asn
885 890 895

Ala Glu Leu Arg Asn Leu Gly Gln Leu Asn Ile Gly Ser Ala Ser Cys
900 905 910

Arg Ile Asn His Ile Leu Asp Thr Val Ser Thr Ile Leu Ser Ala Ile
915 920 925

Cys Gly Ser Ala Ser Asp Gln Ser Asn Pro Gly Ser Val Ser Gln
930 935 940

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:[illegible]:

[illegible]

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

CGATATCTGC AGAATTCAGC TTTAAAGCCC ATTTGCGA 38

(2) INFORMATION FOR SEQ ID NO:206:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

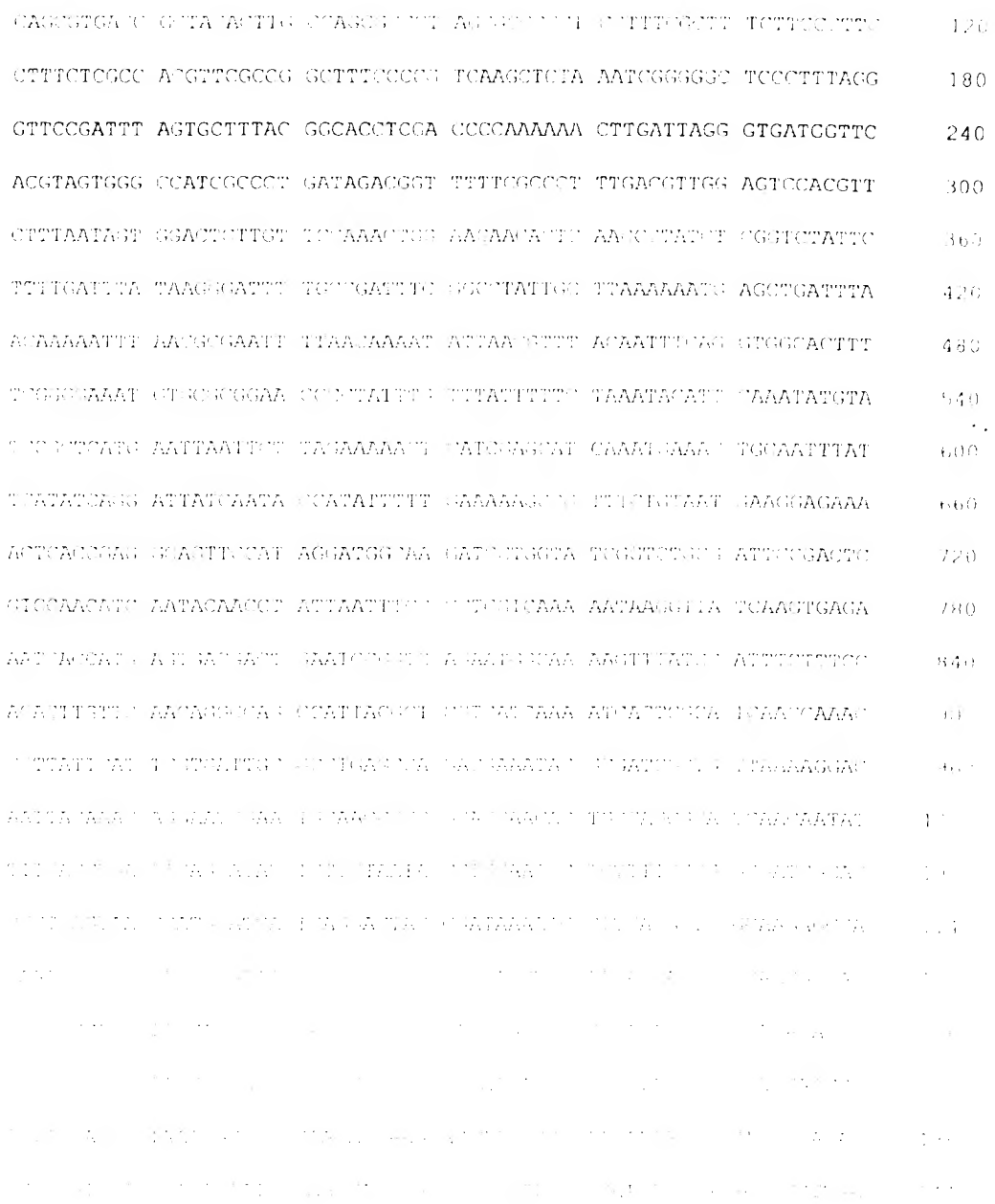
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:[illegible]:

[illegible] 30

(2) INFORMATION FOR SEQ ID NO:[illegible]:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: [illegible] base pairs
(B) TYPE: [illegible]
(C) STRANDEDNESS: [illegible]
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:[illegible]:



AACTCTGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC 1740

AGTGGCGATA AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG 1800

CAGCGGTCGG GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC 1860

ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA 1920

AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT 1980

CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCGGGT TTCGCCACCT CTGACTTGAG 2040

CGTCGATTTT TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG 2100

GCCTTTTTAC GGTTCCTGGC CTTTTGCTGG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA 2160

TCCCCTGATT CTGTGGATAA CCGTATTACC GCCTTTGAGT GAGCTGATAC CGCTCGCCGC 2220

AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG CGGAAGAGCG CCTGATGCGG 2280

TATTTTCTCC TTACGCATCT GTGCGGTATT TCACACCGCA TATATGGTGC ACTCTCAGTA 2340

CAATCTGCTC TGATGCCGCA TAGTTAAGCC AGTATACACT CCGCTATCGC TACGTGACTG 2400

GGTCATGGCT GCGCCCCGAC ACCCGCCAAC ACCCGCTGAC GCGCCCTGAC GGGCTTGTCT 2460

GCTCCCGGCA TCCGCTTACA GACAAGCTGT GACCGTCTCC GGGAGCTGCA TGTGTCAGAG 2520

GTTTTCACCG TCATCACCGA AACGCGCGAG GCAGCTGCGG TAAAGCTCAT CAGCGTGGTC 2580

GTGAAGCGAT TCACAGATGT CTGCCTGTTC ATCCGCGTCC AGCTCGTTGA GTTTCTCCAG 2640

AAGCGTTAAT GTCTGGCTTC TGATAAAGCG GGCCATGTTA AGGGCGGTTT TTTCCTGTTT 2700

GGTCACTGAT GCCTCCGTGT AAGGGGGATT TCTGTTCATG GGGGTAATGA TACCGATGAA 2760

ACGAGAGAGG ATGCTCACGA TACGGGTTAC TGATGATGAA CATGCCCGGT TACTGGAACG 2820

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

CCGGAAGGAG CTGACTGGGT TGAAGGCTCT CAAGGGCATC GGTCGAGATC CCGGTGCCTA	3480
ATGAGTGAGC TAACTTACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA	3540
CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT	3600
[illegible]	3660
[illegible]	3720
[illegible]	3780
[illegible]	3840
[illegible]	3900
[illegible]	3960
[illegible]	4020
AGACAGAACT TAATGGGCCC GCTAACAGCG CGATTTGCTG GTGACCCAAT GCGACCAGAT	4080
[illegible]	4140
[illegible]	4200
[illegible]	4260
[illegible]	432
[illegible]	438
[illegible]	444
[illegible]	450
[illegible]	456
[illegible]	462

AATTAATACG ACTCACTATA GGGGAATTGT GAGCGGATAA CAATTCCCCT CTAGAAATAA 5040

TTTTGTTTAA CTTTAAGAAG GAGATATACA TATGGGCCAT CATCATCATC ATCACGTGAT 5100

CGACATCATC GGGACCAGCC CCACATCCTG GGAACAGCCG GCGGCGGAGG CGGTCCAGCG 5160

GGCGCCGGAT AGCGTCGATG ACATCCGCGT CGCTCGGGTC ATTGAGCAGG ACATGGCCGT 5220

GGACAGCGCC GGCAACATCA CCTACCGCAT CAAGCTCGAA CTGTCGTTTA AGATGAGGCC 5280

GGCGCAACCG AGGGGCTGA AACCACCGAG CGCTTTGCCT GAAACGGGGG CCGGCGCCGG 5340

TAGTGTTGGG ACTACCTCCA CCTCTGGCAC GTTACTCTG GCGGAGACCA GTAGCACCCT 5400

GCTCTACCTG CTGTTCAACC ACGGGGTCGC GTTTTTCAC GAGAGGTATC CGAACGTCAC 5460

GATCACCGCT CAGGGCATCG TTTTCGTCGC TGGATCGCTG CACTCGGTCG CCGGGACGGT 5520

CAACATTGGG GCCTCGACCA CCTATTCTGT GAAAGTCGAT ATTACCGGTA TCAAGGAGGT 5580

GATGAACATC GCCCTAGCGT TCTTCGCTCA CAACTACAAC TATAACCTGC CGCTAGTGAG 5640

CGAGCACCTC AAGTTGAACG GAAAAGTCCT GCCTCCTACT TATCACGGCA GCATCAAAAC 5700

GTGGGATGAT CCTAGATTCC TTGCGCTGAA GCCCATGTAT AACTGGTTCG CACTGCGGGT 5760

AGTTCCCTTC GATGCTTACC AGGATCCGAA TGACTTTC TTTTCACTCA AGTACCTGTC 5820

TAACCAAGAT CCCAAGTCT GGTTAAGTG CTGCATTTT GCACCACCG TCGACTTCCC 5880

CCTCTTCCCC CTTCTCCTAC ACATCGCCG CGCCACCTCT GCTAGCCTCC TTCAAGGCT 5940

GATACCAGAT CTCGAGCTCA AGCTTCGAAT TCTGCAGTCG ACGGTACCGC GGGCCCGGGA 6000

TCCACTAGTA ACGGCCGCCA GTGTGCTGGA ATTCGCCCTT AAGGGCGAAT TCCAGCACAC 6060

GGCGCTGTCG ACTGCTTCAG CTGCAACTTC ACTGCGACA CTTGTTCTTC CCCTACGACC 6720

GGCCGCCGCC AACACGTCGA ATGTCCAGCC GGGCGATCCC AACGCAGCAC CTCCGCCGGC 6780

CGACCCGAAC GCACCGCCGC CACCTGTCAT TGCCCCAAAC GCACCCCAAC CTGTCCGGAT 6840

CGACAACCCG GTTGGAGGAT TCAGCTTCGC GCTGCCTGCT GGCTGGGTGG AGTCTGACGC 6900

CGCCCACTTC GACTACGGTT CAGCACTCCT CAGCAAAACC ACCGGGGACC CGCCACTTCC 6960

CTACAGCGCC CCTCCTCCTC CAATGACAC TTTATCGTCG CTGGCCCCGC TAGACGAAAA 7020

GCTTTACCTC ACCCTCAAG CCACCGACTT CAACGTCTTC GCCGCGTTTG GTCTGGACAT 7080

CGCTGATTTC TATATCCCT ACTCGGCAC CTCAATGAA CTAGAAAACG TCTCCCTTGA 7140

CGTCAACGTC TTCTTGGAA CCGCCTTCA TTACGAATT AACTTCTTCT ATCCGACTAA 7200

TCGGAATGGC CACATCTTGA CGGGCGTAAT CCTTCTGGA CCCCCAACT CACCGACCC 7260

TGCCGACTCT CACGTCGTAT CTGTCGTAAG CTTCGAGAT CCTAACTACT CGCTTGACAA 7320

GCCCGCCCCT AAGCCCTCGT CCGAATCGAT CTCGATTTTC CTCAGCCTCT CCTCCGCCC 7380

GCAACGCCA CTGAGAGCT CCGTCTCCTT CTCCGGGGAA TCGTCTCTAC 7440

CTCGACGACA CCGACTACGG CATCGACTCT ACCTCCTTCA GAATTTTCA GATATCCATC 7500

ACACTGACCT CCGCCTGAGC ACCACGACTA CATCACTTCA GATCCTGCCT CTAACAAAGC 7560

CCAAAAAAA CCTGATTTCC CTCATTAAC CTGACTACTT CATCCCCGGC CAGCGTGGTA 7620

CCTTTTAAA CTCCTTGCAA CTCTTTCAT CCCCTCCATC CCCCAAAAT TCAGAATAAT 7680

Asp Ser Val Asp Asp Ile Arg Val Ala Arg Val Ile Glu Gln Asp Met
45 50 55

Ala Val Asp Ser Ala Gly Lys Ile Thr Tyr Arg Ile Lys Leu Glu Val
50 55 60

Ser Phe Lys Met Arg Pro Ala Gln Pro Arg Gly Ser Lys Pro Pro Ser
65 70 75 80

Gly Ser Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Asn Thr Thr Pro
85 90 95

Ala Ser Ser Pro Val Thr Leu Ala Glu Thr Gly Ser Thr Leu Leu Tyr
100 105 110

Pro Leu Phe Asn Leu Trp Gly Pro Ala Phe His Gly Arg Tyr Pro Asn
115 120 125

Val Thr Ile Thr Ala Gln Gly Leu Gly Ser Gly Ala Gly Ile Ala Gln
130 135 140

Ala Ala Ala Gly Thr Val Asn Ile Gly Ala Ser Asn Ala Tyr Leu Ser
145 150 155 160

Glu Gly Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala
165 170 175

Ile Ser Ala Gln Gln Val Asn Tyr Arg Leu Pro Gly Val Ser His His
180 185 190

Leu Leu Leu Asn Gly Lys Thr Leu Ala Asn His Gly Gly Thr Ile
195 200 205

Tyr Thr Trp Thr Asp Pro Gln Ile Ala Ile Asn Gln Arg
210 215

Leu Ser Ala Val Val Ile Leu Asp Asp Leu Gly
220 225

Gly Phe Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp
340 345 350

Gly Pro Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile
355 360 365

Val Asn Asn Arg Glu Lys Asp Ala Ala Thr Asn Gln Thr Leu Gln Ala
370 375 380

Ile Leu Met Trp Ala Ile Thr Asp Gly Asn Lys Asn Ser Phe Leu Asp
385 390 395 400

Glu Val His Phe Glu Pro Leu Pro Phe Ala Val Val Lys Leu Ser Asp
405 410 415

Ala Leu Glu Ala Thr Glu Ser Ala Leu Met Lys Thr Asp Ala Ala
420 425 430

Thr Leu Ala Lys Val Ala Ala Gly Asn Phe Ser Asn Ile Ser Gly Asp Leu
435 440 445

Lys Thr Glu Ile Asp Glu Glu Ser Thr Ala Gly Ser Leu Gln Gly
450 455 460

Ala Pro Arg Gly Ala Ala Gly Thr Asn Ala Asn Ala Val Val Arg
465 470 475 480

Ile Leu Leu Arg Asn Lys Gln Leu Leu Val Lys Ser
485 490

Xxx Asp Xxx Xxx Ile Xxx Gly Xxx Xxx Xxx Pro Val Asn Asn Asp Thr
625 630 635 640

Arg Ile Val Leu Gly Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu
645 650 655

Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu
660 665 670

Ile Tyr Met Xxx Tyr Xxx Gly Thr Arg Ile Asn Gln Glu Thr Val Ser
675 680 685

Leu Asp Ala Asn Gly Val Ser Gly Ser Ala Xxx Xxx Tyr Xxx Glu Val Lys
690 695 700

Ile Ser Asp Xxx Ser Xxx Xxx Xxx Asn Gly Xxx Ile Xxx Trp Xxx Gly Val Ile
705 710 715 720

Gly Xxx Xxx Ala Ala Asn Ala Xxx Asn Ala Xxx Gly Xxx Xxx Ala Xxx Trp
725 730 735

Ile Val Val Trp Leu Gly Thr Asn Asn Asn Xxx Val Asp Lys Gly Ala
740 745 750

Asn Lys Asn Xxx Ala Glu Xxx Xxx Asn Xxx Xxx Leu Val Xxx Xxx Xxx Xxx
755 760 765

Ala Xxx Asn Xxx Ala Xxx Arg Xxx Xxx Asn Xxx Xxx Ala Xxx Ala
770 775 780

Gly Leu Val Xxx Xxx Ile Xxx Xxx Xxx Xxx Xxx Xxx Arg Xxx Ile
785 790 795 800

CLAIMS

We claim:

1. A polypeptide comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

(a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu (SEQ ID NO: 115);

(b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala Leu Gly-Thr-Pro-Ala-Pro-Ser (SEQ ID NO: 116);

(c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg (SEQ ID NO: 17);

(d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro (SEQ ID NO: 118);

(e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val (SEQ ID NO: 119);

(f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID NO: 120);

(g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser (SEQ ID NO: 121);

(h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly (SEQ ID NO: 122);

(i) Asp Pro Ala Ser-Ala Pro Asp Val Pro-Thr-Ala-Ala-Gln-Leu Thr-Ser Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn Val Ser Phe Ala-Asn (SEQ ID NO: 123); and

(j) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID NO: 131)

wherein Xaa may be any amino acid

2. A polypeptide comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

(a) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID NO: 124) and

(b) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID NO: 132), wherein Xaa may be any amino acid.

3. A polypeptide comprising an antigenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 1, 2, 4-10, 13-25, 52, 94 and 96 or a complement thereof under moderately stringent conditions.

4. A polypeptide comprising an antigenic portion of a *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NOS: 26-51, 133, 134, 158-178 and 196, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 26-51, 133, 134, 158-178 and 196 or a complement thereof under moderately stringent conditions.

5. A DNA molecule comprising a nucleotide sequence encoding a polypeptide according to any one of claims 1-4

6. A recombinant expression vector comprising a DNA molecule according to claim 5.

7. A host cell transformed with an expression vector according to claim 6.

8. The host cell of claim 7 wherein the host cell is selected from the group consisting of *E. coli*, yeast and mammalian cells.

9. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting a biological sample with one or more polypeptides according to any of claims 1-4; and

(b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.

10. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting a biological sample with a polypeptide having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and

(b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.

11. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting a biological sample with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198, the complements of said sequences, and DNA sequences

(b) detecting in the sample the presence of antibodies that bind to at least one of the polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample.

12. The method of any one of claims 9-11 wherein step (a) additionally comprises contacting the biological sample with a 38 kD *M. tuberculosis* antigen and step (b) additionally comprises detecting in the sample the presence of antibodies that bind to the 38 kD *M. tuberculosis* antigen.

13. The method of any one of claims 9-11 wherein the polypeptide(s) are bound to a solid support.

14. The method of claim 13 wherein the solid support comprises nitrocellulose, latex or a plastic material.

15. The method of any one of claims 9-11 wherein the biological sample is selected from the group consisting of whole blood, serum, plasma, saliva, cerebrospinal fluid and urine.

16. The method of claim 15 wherein the biological sample is whole blood or serum.

17. A method for detecting *M tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with at least two oligonucleotide primers in a polymerase chain reaction, wherein at least one of the oligonucleotide primers is specific for a DNA molecule according to claim 5, and

(b) detecting in the sample a DNA sequence that amplifies in the presence of the oligonucleotide primers, thereby detecting *M. tuberculosis* infection.

18. The method of claim 17, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotides of a DNA molecule according to claim 5.

19. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with at least two oligonucleotide primers in a polymerase chain reaction, wherein at least one of the oligonucleotide primers is specific for a DNA sequence selected from the group consisting of SEQ ID NOS. 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and

(b) detecting in the sample a DNA sequence that amplifies in the presence of the first and second oligonucleotide primers, thereby detecting *M. tuberculosis* infection.

20. The method of claim 19, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

21. The method of claims 17 or 19 wherein the biological sample is selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

22. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with one or more oligonucleotide probes specific for a DNA molecule according to claim 5; and

(b) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe, thereby detecting *M. tuberculosis* infection.

23. The method of claim 22 wherein the probe comprises at least about 15 contiguous nucleotides of a DNA molecule according to claim 5.

24. A method for detecting *M. tuberculosis* infection in a biological sample, comprising:

(a) contacting the sample with one or more oligonucleotide probes specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and

(b) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe, thereby detecting *M. tuberculosis* infection.

25. The method of claim 24 wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

26. The method of claims 22 or 24 wherein the biological sample is selected from the group consisting of whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine.

27. A method for detecting *M tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide according to any one of claims 1-4; and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

28. A method for detecting *M tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide having an N-terminal sequence selected from the group consisting of sequences provided in SEQ ID NO: 129 and 130; and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

29. A method for detecting *M tuberculosis* infection in a biological sample, comprising:

(a) contacting the biological sample with a binding agent which is capable of binding to a polypeptide encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and

(b) detecting in the sample a protein or polypeptide that binds to the binding agent, thereby detecting *M. tuberculosis* infection in the biological sample.

30. The method of any one of claims 27-29 wherein the binding agent is a monoclonal antibody.

31. The method of any one of claims 27-29 wherein the binding agent is a polyclonal antibody.

32. A diagnostic kit comprising:

(a) one or more polypeptides according to any of claims 1-4; and

(b) a detection reagent.

33. A diagnostic kit comprising:

(a) one or more polypeptides having an N-terminal sequence selected from

34. A diagnostic kit comprising:

(a) one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198; and

(b) a detection reagent.

35. The kit of any one of claims 32-34 wherein the polypeptide(s) are immobilized on a solid support.

36. The kit of claim 35 wherein the solid support comprises nitrocellulose, latex or a plastic material.

37. The kit of any one of claims 32-34 wherein the detection reagent comprises a reporter group conjugated to a binding agent.

38. The kit of claim 37 wherein the binding agent is selected from the group consisting of anti-immunoglobulins, Protein G, Protein A and lectins.

39. The kit of claim 37 wherein the reporter group is selected from the group consisting of radioisotopes, fluorescent groups, luminescent groups, enzymes, biotin and dye particles.

40. A diagnostic kit comprising at least two oligonucleotide primers, at least one of the oligonucleotide primers being specific for a DNA molecule according to claim 5.

41. A diagnostic kit according to claim 40, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotide of a DNA molecule according to claim 5.

42. A diagnostic kit comprising a at least two oligonucleotide primers, at least one of the primers being specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

43. A diagnostic kit according to claim 42, wherein at least one of the oligonucleotide primers comprises at least about 10 contiguous nucleotide of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

44. A diagnostic kit comprising at least one oligonucleotide probe, the oligonucleotide probe being specific for a DNA molecule according to claim 5.

45. A kit according to claim 44, wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA molecule according to claim 5.

46. A diagnostic kit comprising at least one oligonucleotide probe, the oligonucleotide probe being specific for a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

47. A kit according to claim 46, wherein the oligonucleotide probe comprises at least about 15 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NOS: 3, 11, 12, 135, 136, 151-155, 184-188, 194-195 and 198.

48. A monoclonal antibody that binds to a polypeptide according to any of

49. A polyclonal antibody that binds to a polypeptide according to any of claims 1-4.

50. A fusion protein comprising two or more polypeptides according to any one of claims 1-4.

51. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and ESAT-6 (SEQ ID NO: 99).

52. A fusion protein comprising a polypeptide having an N-terminal sequence selected from the group of sequences provided in SEQ ID NOS: 129 and 130.

53. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and the *M. tuberculosis* antigen 38 kD (SEQ ID NO: 150).

54. A diagnostic kit comprising:

(a) one or more fusion proteins according to any one of claims 50-53; and

(b) a detection reagent.

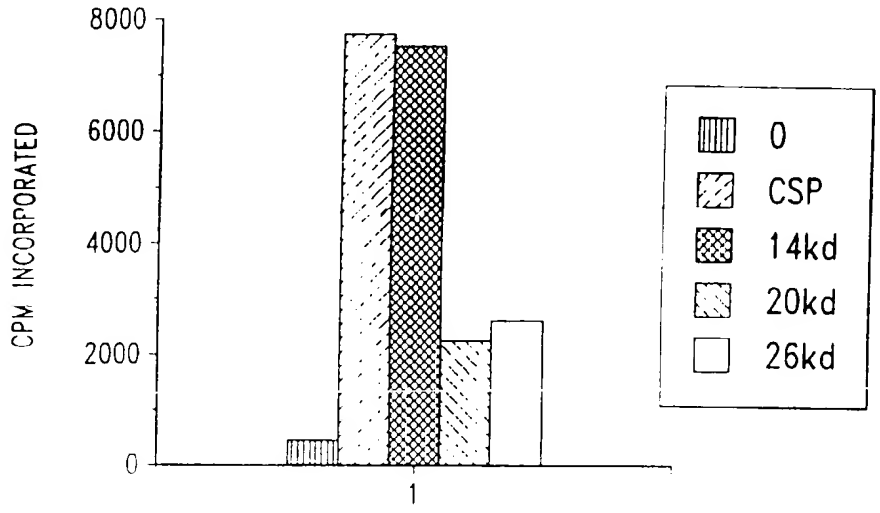


Fig. 1A-1

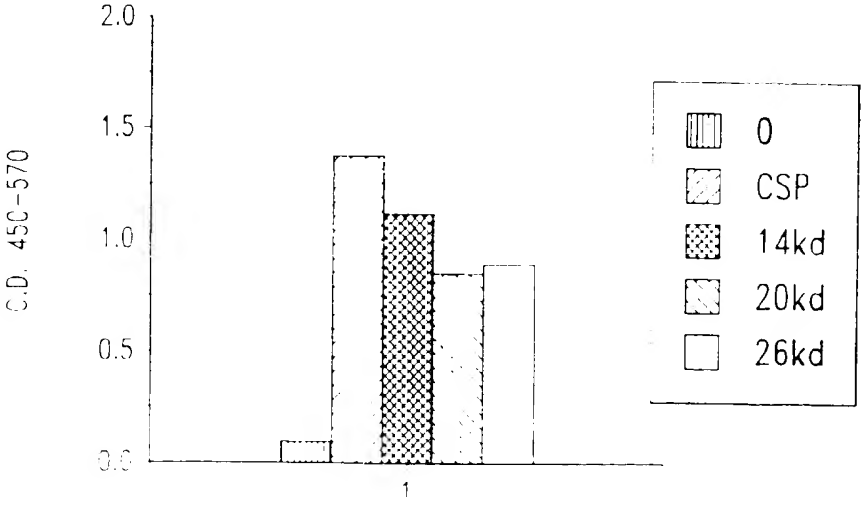


Fig. 1A-2

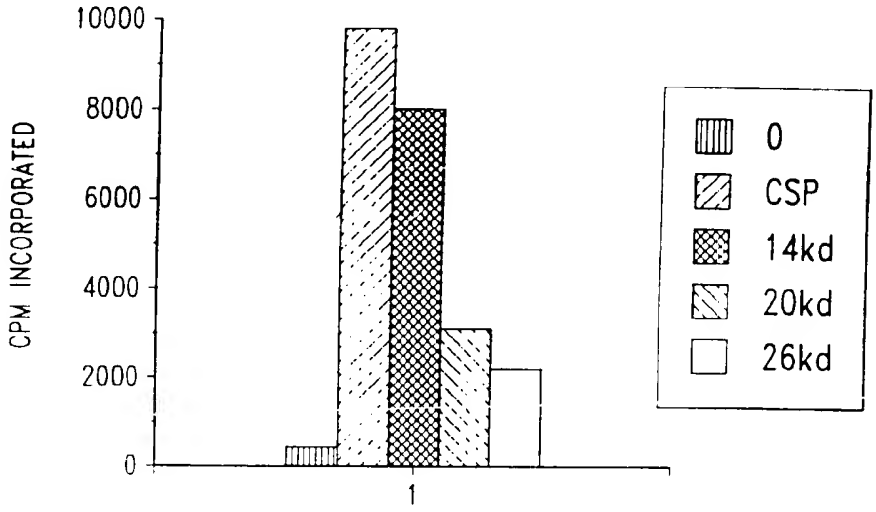

10000
8000
6000
4000
2000
0
CPM INCORPORATED
1
0
CSP
14kd
20kd
26kd

Fig. 1B-1

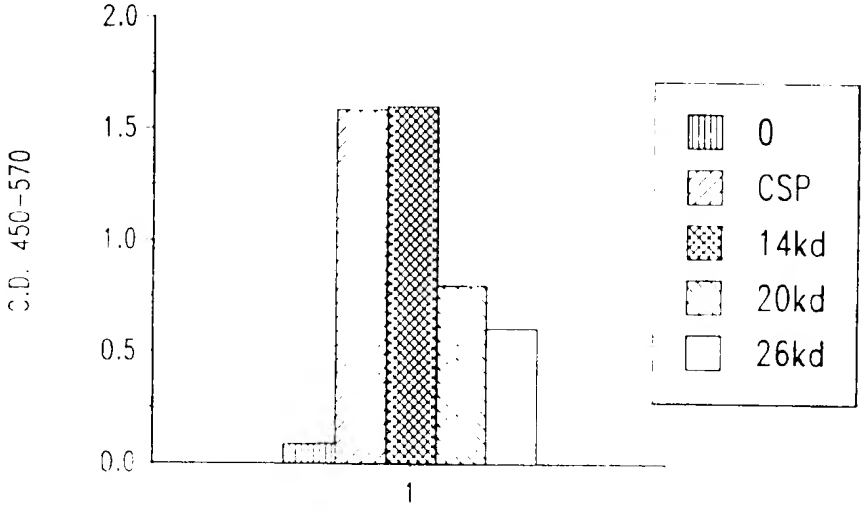
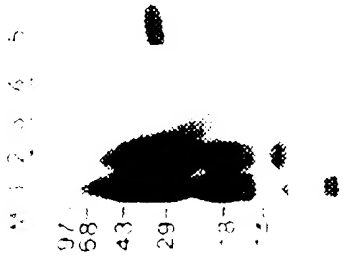
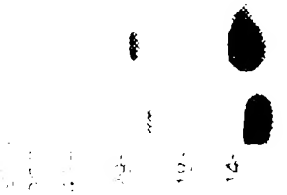

2.0
1.5
1.0
0.5
0.0
O.D. 450-570
1
0
CSP
14kd
20kd
26kd

Fig. 2B

Fig. 2D





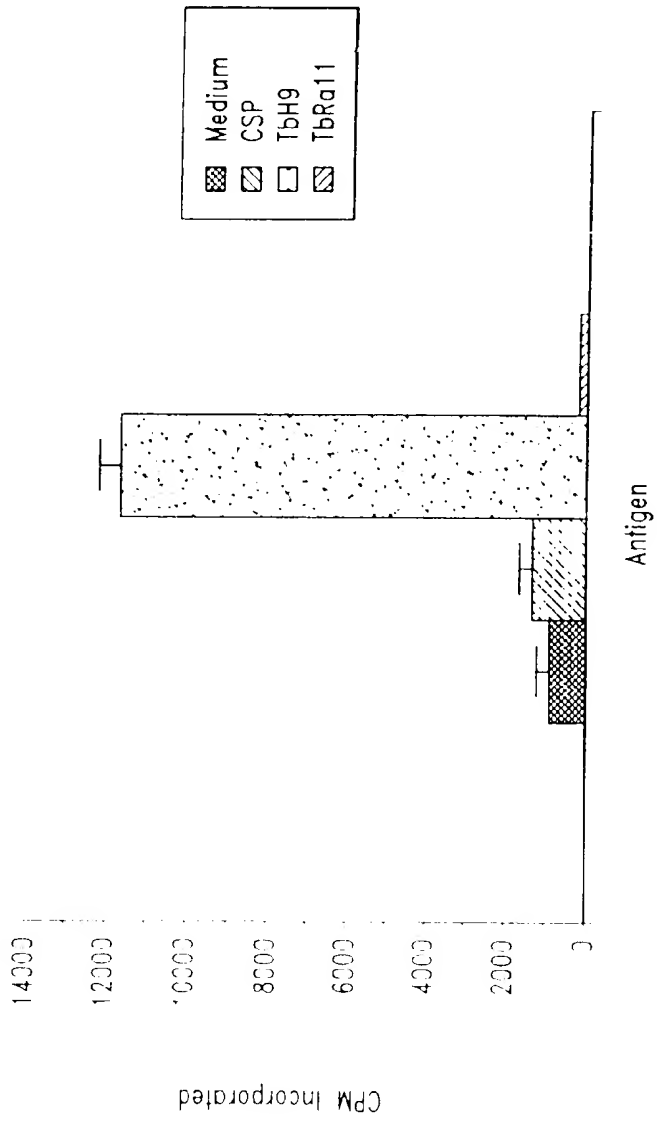

Medium
CSP
TbH9
TbRa11
Antigen
CPM Incorporated
14000
12000
10000
8000
6000
4000
2000
0

Fig. 3A

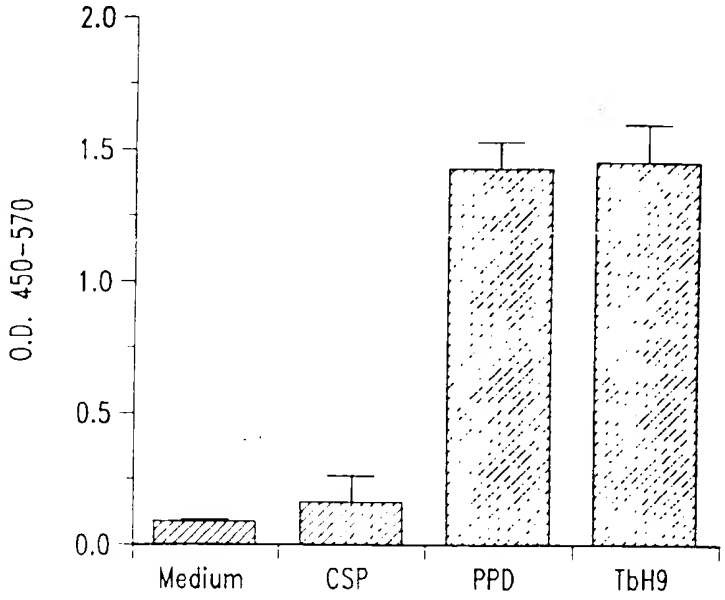


Fig. 3B

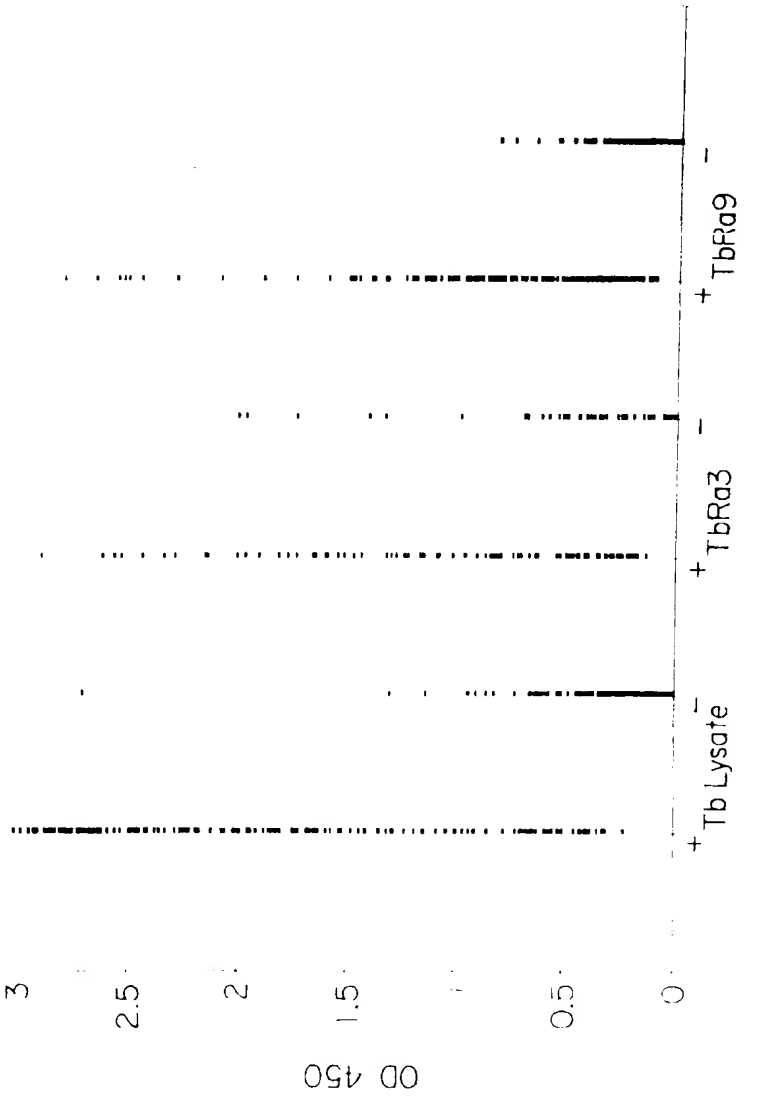


Fig. 4

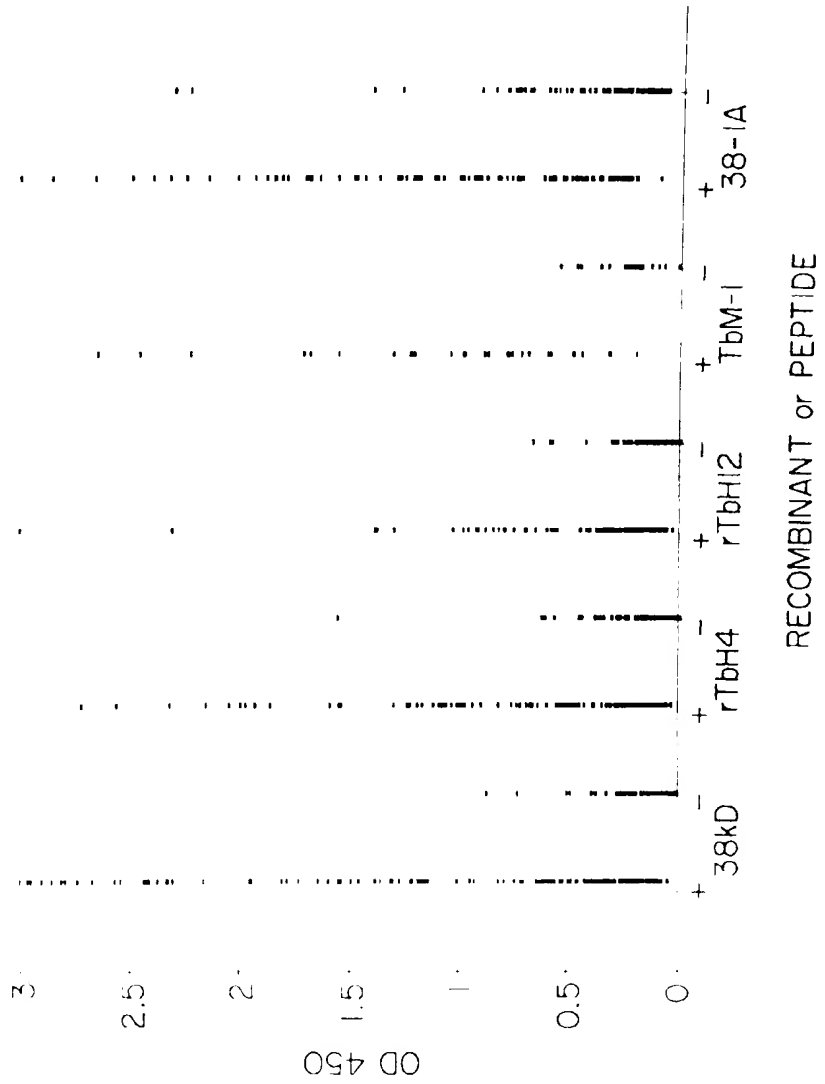

OD 450
3
2.5
2
1.5
1
0.5
0
38kD
rTbH4
rTbH12
TbM-1
38-1A
RECOMBINANT or PEPTIDE

Fig. 5

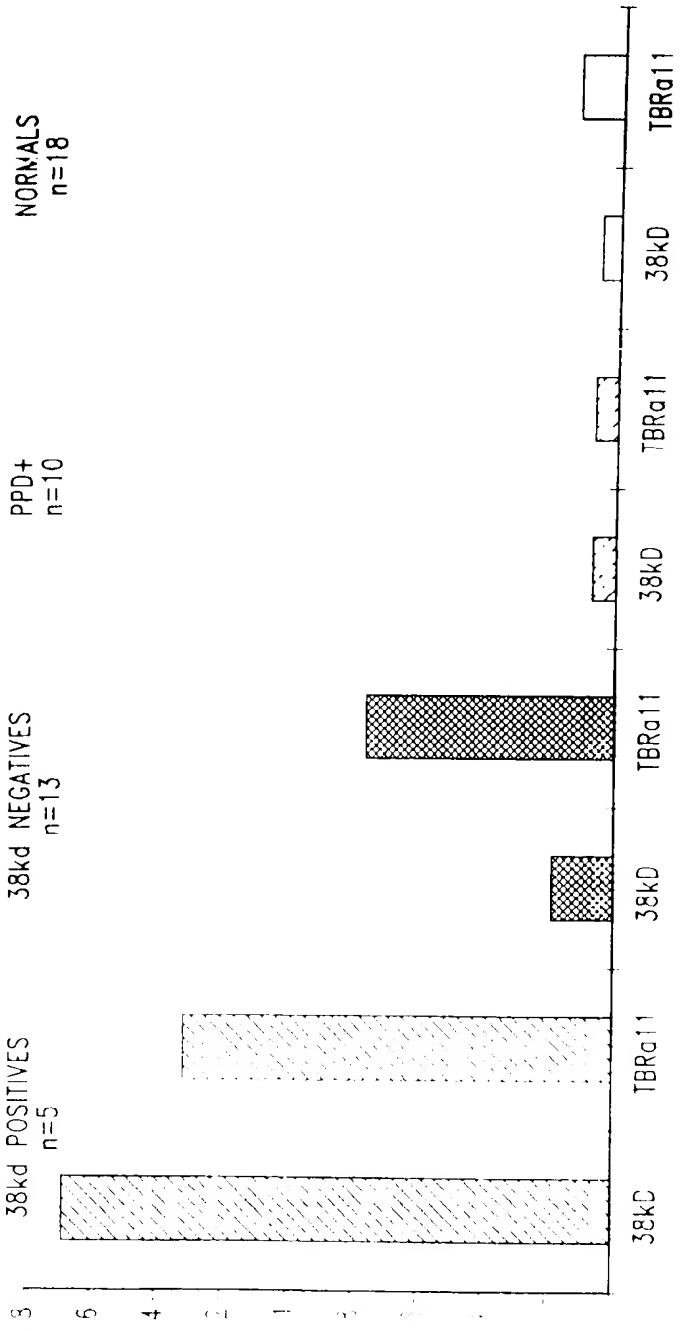

38kd POSITIVES n=5
38kd NEGATIVES n=13
PPD+ n=10
NORMALS n=18
38kD
TBRa11
38kD
TBRa11
38kD
TBRa11
38kD
TBRa11
RECOMBINANT

Fig. 6.

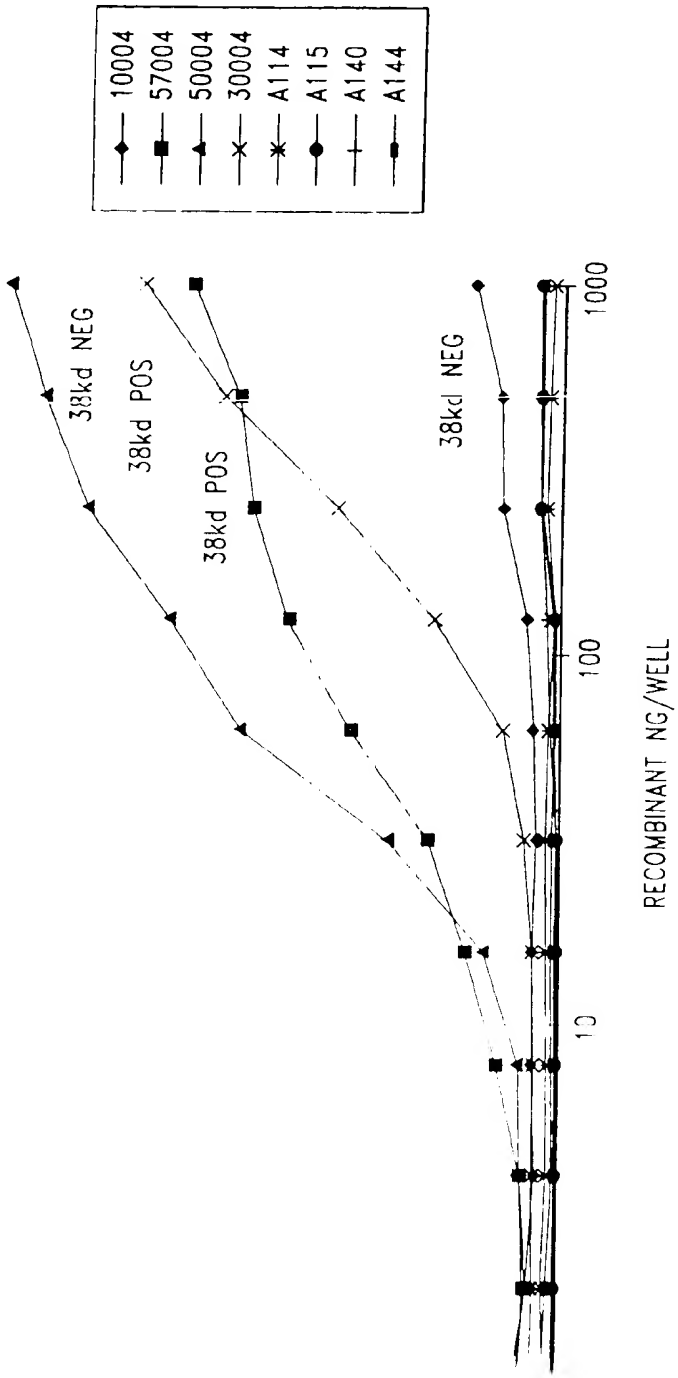

10004
57004
50004
30004
A114
A115
A140
A144
38kd NEG
38kd POS
38kd POS
38kd NEG
1000
100
10
RECOMBINANT NG/WELL

Fig. 7

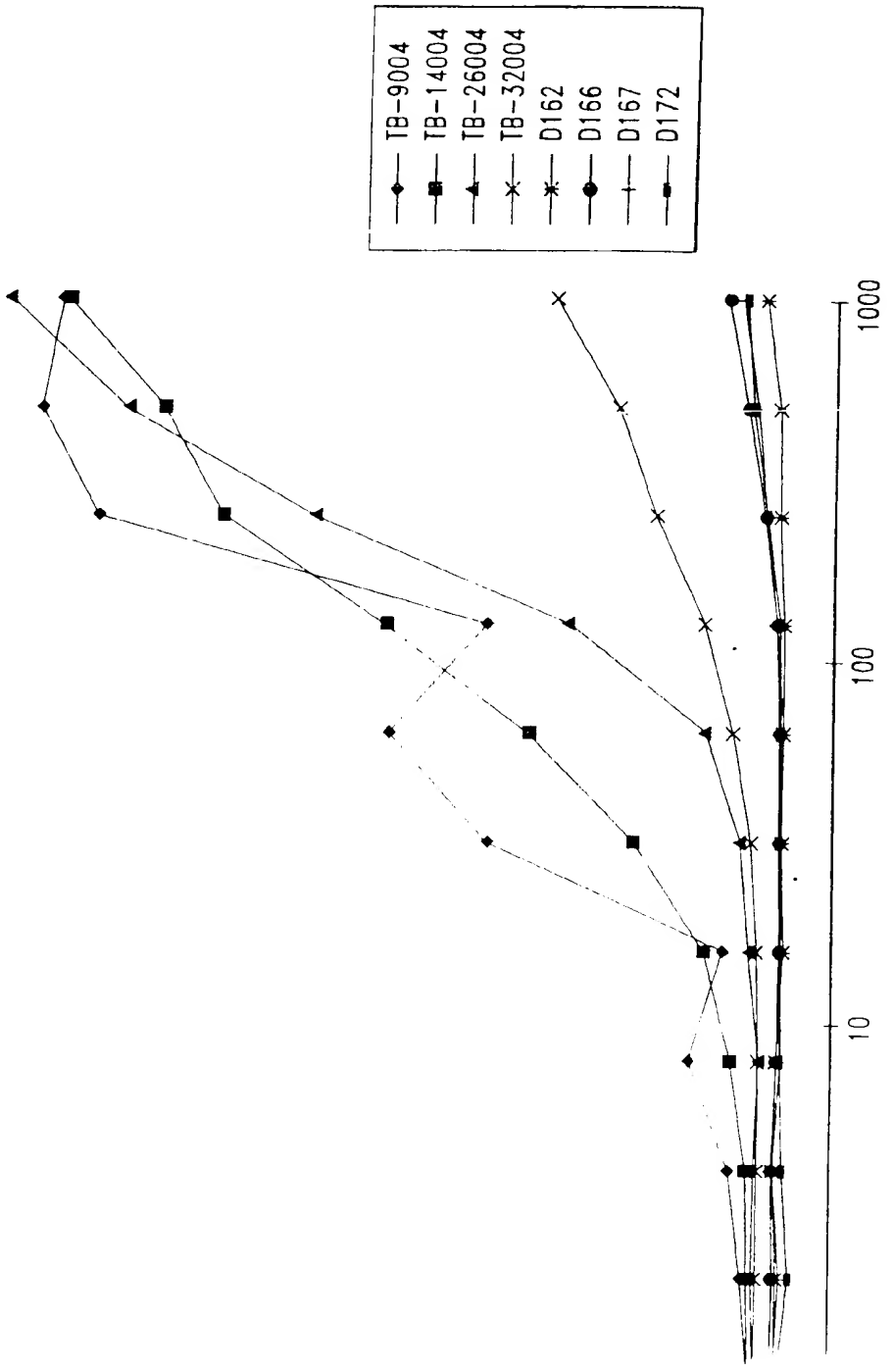

TB-9004
TB-14004
TB-26004
TB-32004
D162
D166
D167
D172
1000
100
10
RECOMBINANT NG/WELL

Fig. 8

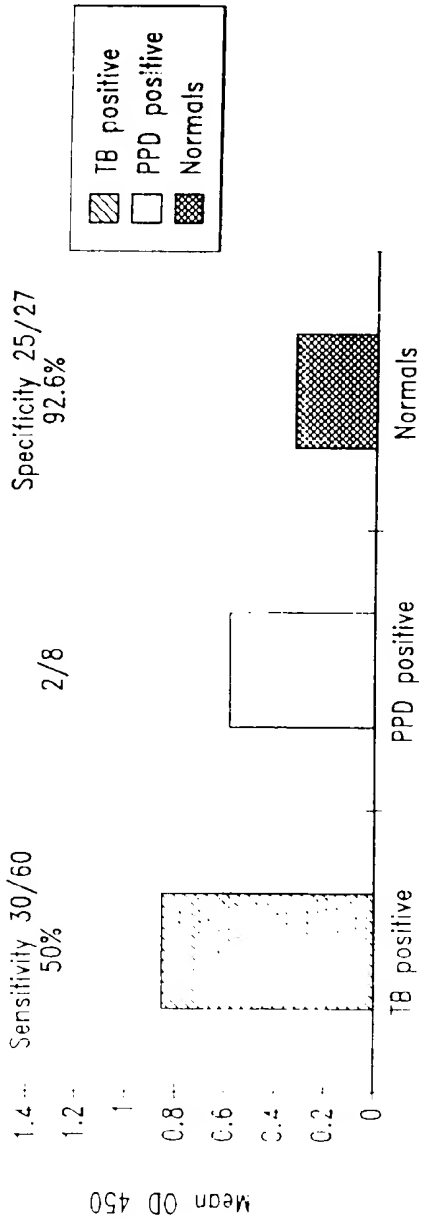

TB positive
PPD positive
Normals
Specificity 25/27
92.6%
2/8
Sensitivity 30/60
50%
Normals
PPD positive
TB positive
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Mean OD 450

Fig. 9

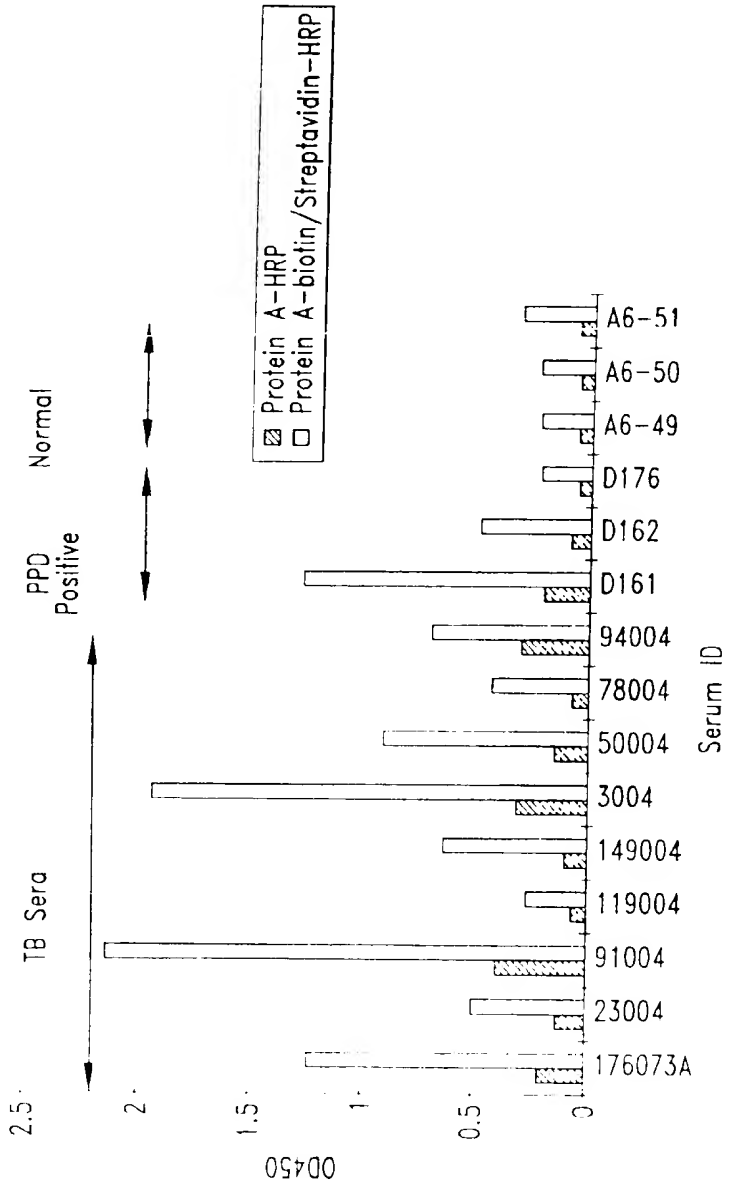

Protein A-HRP
Protein A-biotin/Streptavidin-HRP
Normal
PPD Positive
TB Sera
A6-51
A6-50
A6-49
D176
D162
D161
94004
78004
50004
3004
149004
119004
91004
23004
176073A
Serum ID
2.5
2
1.5
1
0.5
0
OD450

Fig. 10

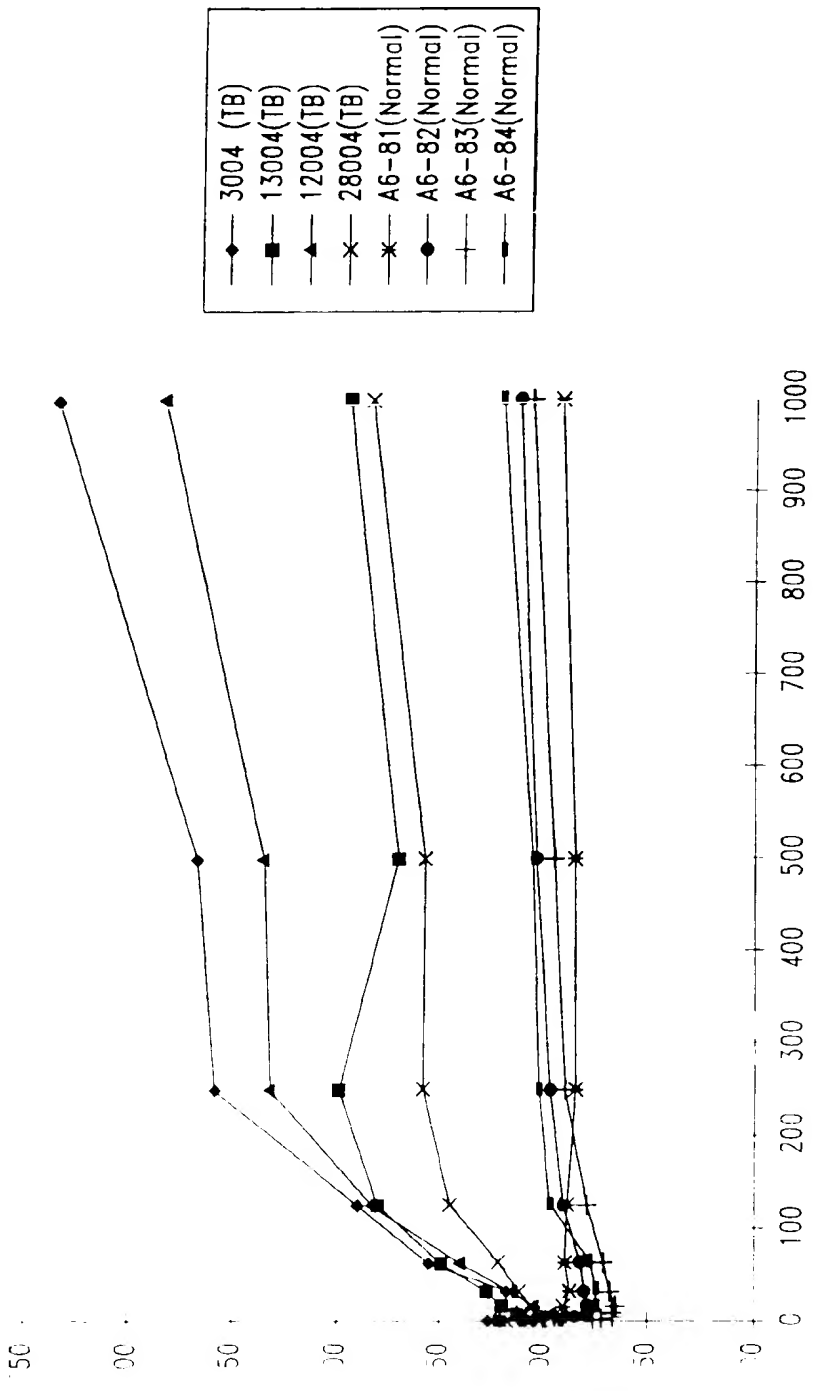

3004 (TB)
13004(TB)
12004(TB)
28004(TB)
A6-81(Normal)
A6-82(Normal)
A6-83(Normal)
A6-84(Normal)
0
100
200
300
400
500
600
700
800
900
1000
Recombinant ng/well

Fig. 11